Quantitative structure-retention relationships for rapid method development in hydrophilic interaction liquid chromatography of pharmaceutical compounds

Maryam Taraji PhD

School of Physical Sciences
University of Tasmania
March 2017

Human being are members of a whole,

In creation of one essence and soul.

If one member is afflicted with pain,

Other members' unease will remain.

If you have no sympathy for human pain,

The name of human you cannot retain.

(Saadi of Shiraz, Iran)

Declaration

This thesis contains no material which has been accepted for a degree or diploma by the University or any other institution, except by way of background information and duly acknowledged in the thesis, and to the best of the my knowledge and belief no material previously published or written by another person except where due acknowledgement is made in the text of the thesis, nor does the thesis contain any material that infringes copyright.

The publishers of the papers in this thesis hold the copyright for that content, and access to the material should be sought from the respective journals. The remaining non published content of the thesis may be made available for loan and limited copying and communication in accordance with the Copyright Act 1968.

The research associated with this thesis abides by the international and Australian codes on human and animal experimentation, the guidelines by the Australian Government's Office of the Gene Technology Regulator and the rulings of the Safety, Ethics and Institutional Biosafety Committees of the University

Maryam Taraji
March 2017

Acknowledgements

I would like to sincerely thank my primary supervisor, Prof. Paul Haddad, for his guidance and support throughout this study. I have learned so much, and without you, this would not have been possible. Thank you so much for making this journey such a great experience for me!

I would also like to thank my co-supervisors, Dr. Ruth Amos and Assoc/Prof. Robert Shellie, for their excellent guidance and advice. I want to thank Dr. Roman Szucs for his encouragement and help throughout my PhD.

I would like to acknowledge the Tasmanian Partnership for Advanced Computing (TPAC) for allocation of computing resources. I also acknowledge the Australian Research Council for the financial support of this research by an ARC Linkage Projects grant (LP120200700) and the Australian Commonwealth Government for providing me an International Postgraduate Research Scholarship (IPRS).

To my real friends from Australia, Julie Viecieli, Max Bingham, and Geoff Howard, thank you for your understanding and encouragement in my many moments of crisis. Your friendship makes my life a wonderful experience. Also to the students, post docs and staff in ACROSS and the Department of Chemistry for offering their assistance and friendship throughout my candidature.

Last but surely not least, I am thankful to my family for their love, supporting my dreams and always believing in me. To my beloved husband, Amin, thank you for supporting me for everything.

Statement of co-authorship

The following people and institutions contributed to the publication of work undertaken as part of this thesis:

Maryam Taraji, School of Physical Sciences, UTAS = Candidate

Paul R. Haddad, School of Physical Sciences, UTAS = Author 1

Ruth Amos, School of Physical Sciences, UTAS = Author 2

Mohammad Talebi, School of Physical Sciences, UTAS = Author 3

Roman Szucs, Pfizer Global Research and Development, Sandwich, United Kingdom = Author 4

John W. Dolan, LC Resources, McMinnville, Oregon, United States = Author 5

Chris A. Pohl, Thermo Fisher Scientific, Sunnyvale, California, United States = Author 6

Author details and their roles:

Paper 1 < Prediction of retention in hydrophilic interaction liquid chromatography using solute molecular descriptors based on chemical structures > Located in ***chapter 3***

Candidate was the primary author (50 %) with authors 1 (20 %) and 2 (10 %) contributed to the idea, and its formalisation, development. Authors 3, 4, 5, and 6 assisted with modelling tools and instrument maintenance (5 % respectively).

Paper 2 < Rapid method development in hydrophilic interaction liquid chromatography for pharmaceutical analysis using a combination of quantitative structure-retention relationships and design of experiments > Located in ***chapter 4***

Candidate was the primary author (50 %) with authors 1 (20 %) and 2 (10 %) contributed to the idea, and its formalisation, development. Authors 3, 4, 5, and 6 assisted with modelling tools and instrument maintenance (5 % respectively).

Paper 3 < Use of dual-filtering to create training sets leading to improved accuracy in quantitative structure-retention relationship modelling for hydrophilic interaction liquid chromatographic systems > Located in ***chapter 5***

Candidate was the primary author (50 %) with authors 1 (20 %) and 2 (10 %) contributed to the idea, and its formalisation, development. Authors 3, 4, 5, and 6 assisted with modelling tools and instrument maintenance (5 % respectively).

We the undersigned agree with the above stated "proportion of work undertaken" for each of the above published (or submitted) peer-reviewed manuscripts contributing to this thesis:

Signed: _____

Paul Haddad
Supervisor
School of Physical Sciences
University of Tasmania

John Dickey
Head of School
School of Physical Sciences
University of Tasmania

List of publications and presentations

1. Maryam Taraji, Paul R. Haddad, Ruth Amos, Mohammad Talebi, Roman Szucs, John W. Dolan, Chris A. Pohl. "Prediction of retention in hydrophilic interaction liquid chromatography using solute molecular descriptors based on chemical structures", Journal of Chromatography A, 1486 (2016) 59-67. (***Chapter 3***)

2. Maryam Taraji, Paul R. Haddad, Ruth Amos, Mohammad Talebi, Roman Szucs, John W. Dolan, Chris A. Pohl. "Rapid method development in hydrophilic interaction liquid chromatography for pharmaceutical analysis using a combination of quantitative structure-retention relationships and design of experiments", Analytical Chemistry, 89 (3) (2017) 1870-1878. (***Chapter 4***)

3. Maryam Taraji, Paul R. Haddad, Ruth Amos, Mohammad Talebi, Roman Szucs, John W. Dolan, Chris A. Pohl. "Use of dual-filtering to create training sets leading to improved accuracy in quantitative structure-retention relationship modelling for hydrophilic interaction liquid chromatographic systems", prepared for Journal of Chromatography A. (***Chapter 5***)

4. Mohammad Talebi, Soo Hyun Park, Maryam Taraji, Yabin Wen, Ruth Amos, Paul R. Haddad, Robert A. Shellie, Roman Szucs, John W. Dolan, Chris A. Pohl. "Retention time prediction based on molecular structure in pharmaceutical method development: A perspective" LCGC, 34 (8) (2016) 550-558.

5. Eva Tyteca, Mohammad Talebi, Ruth Amos, Soo Hyun Park, Maryam Taraji, Yabin Wen, Roman Szucs, Chris A. Pohl, John W. Dolan, Paul R. Haddad "Towards a chromatographic similarity index to establish localized Quantitative Structure-Retention Models for retention prediction: use of retention factor ratio", Journal of Chromatography A, 1486 (2016) 50-58. (***Chapter 5***)

6. Paul R. Haddad, Maryam Taraji, Mohammad Talebi, Robert A. Shellie, Roman Szucs, John W. Dolan, Chris A. Pohl. "Method development in HILIC using quantitative structure-retention relationships based on analyte molecular structures", HPLC 2017, Prague, Czech Republi.

7. Maryam Taraji, Paul R. Haddad, Ruth Amos, Mohammad Talebi, Robert A. Shellie, Roman Szucs, John W. Dolan and Chris A. Pohl. "A quality-by-design methodology for rapid method development in pharmaceutical analysis", ASASS2 2016, Hobart, Australia.

8. Paul R. Haddad, Soo Hyun Park, Maryam Taraji, Yabin Wen, Eva Tyteca, Mohammad Talebi, Ruth Amos, Robert A. Shellie, Roman Szucs, Chris A. Pohl, John W. Dolan "Prediction of chromatographic retention times based on chemical structures of analytes", ASASS2 2016, Hobart, Australia.

9. Maryam Taraji, Paul R. Haddad, Mohammad Talebi, Ruth Amos, Robert A. Shellie, Roman Szucs, John W. Dolan, Chris A. Pohl. "A Quality-by-Design Methodology for Rapid HILIC Method Development in Pharmaceutical Analysis", HPLC 2016, San Francisco, USA.

10. Paul Haddad, Soo Park, Maryam Taraji, Yabin Wen, Mohammad Talebi, Ruth Amos, Robert Shellie, Roman Szucs, John Dolan, Chris Pohl. "Role of Structural Similarity in Prediction of Retention in Reversed-Phase, Ion-Exchange and Hydrophilic Interaction Liquid Chromatography Modes Using Quantitative Structure-Retention Relationships", HPLC 2016, San Francisco, USA.

11. Paul R. Haddad, Soo Hyun Park, Maryam Taraji, Yabin Wen, Eva Tyteca, Mohammad Talebi, Ruth Amos, Robert A. Shellie, Roman Szucs, Chris A. Pohl, John W. Dolan "Prediction of retention times in reversed-phase, ion-exchange and HILIC modes based on chemical structures", ISC 2016, Cork, Ireland.

12. Maryam Taraji, Georg Schuster, Mohammad Talebi, Greg W. Dicinoski, Robert A. Shellie, Paul R. Haddad, Roman Szucs, John W. Dolan, Chris A. Pohl. "HILIC method development in pharmaceutical analysis", HPLC 2015, Geneva, Switzerland.

13. Roman Szucs, Melissa Hanna-Brown, Paul R. Haddad, Soo Hyun Park, Maryam Taraji, Mohammad Talebi. "The role of quantitative structure retention relationships in quality by design chromatographic method development: optimization of retention models", HPLC 2015, Geneva, Switzerland.

14. Paul R. Haddad, Soo Hyun Park, Maryam Taraji, Robert A. Shellie, Greg W. Dicinoski, Georg Schuster, Mohammad Talebi, Roman Szucs, Chris

A. Pohl, John W. Dolan. "Prediction of retention times in reversed-phase, HILIC and ion chromatography based on chemical structures of analytes", HPLC 2015, Beijing, China.

15. Paul R. Haddad, Soo Hyun Park, Maryam Taraji, Robert A. Shellie, Greg W. Dicinoski, Georg Schuster, Mohammad Talebi, Roman Szucs, Chris A. Pohl, John W. Dolan. "Use of quantitative structure-retention relationships to choose the optimal chromatographic techniques", Pacifichem 2015, Honolulu, USA.

16. Maryam Taraji, Mohammad Talebi, George Schuster, Greg W. Dicinoski, Robert A. Shellie, Paul R. Haddad, Roman Szucs, John W. Dolan, Chris A. Pohl. "The role of similarity in QSRR studies", RACI National Congress 2014, Adelaide, Australia.

17. Maryam Taraji, Robert A. Shellie, Greg W. Dicinoski, Mohammad Talebi, Paul R. Haddad, Roman Szucs, John W. Dolan and Chris A. Pohl. "Method development in pharmaceutical analysis: assessment of column selectivity in hydrophilic interaction liquid chromatography using improved quantitative structure–retention relationships models", HPLC 2013, Hobart, Australia.

Abstract

The development of computer-assisted approaches capable of accurate prediction of the retention behaviour of analytes, leading to optimisation of chromatographic performance, is a major goal for method development in chromatography. Statistically-derived quantitative structure-retention relationships (QSRRs) represent a quite popular approach to retention prediction. Hydrophilic interaction chromatography (HILIC) is nowadays well known as a powerful technique for the separation of polar compounds. However, the detailed retention mechanism applicable in HILIC is still under some discussion and for this reason, method development in HILIC is difficult.

The first part of this thesis concerns the application of QSRR methodology to predict the retention times of pharmaceutical test analytes on five HILIC stationary phases (bare silica, amine, amide, diol and zwitterionic), with a view to selecting the most suitable stationary phase(s) for the separation of these analytes. QSRR methodology seeks mathematical equations that correlate molecular features to chromatographic parameters. Genetic algorithm (GA) feature selection and partial least squares (PLS) regression were used to correlate experimental retention data to various density-functional-theory-computed molecular descriptors. The predictive power of the QSRR models was successfully evaluated performing an external validation to predict retention times of test compounds. The QSRR models developed were also utilised to provide some insight into the separation mechanisms operating in the HILIC mode.

The second part of this thesis describes a Quality-by-Design workflow, which combines QSRR methodology with design of experiments (DoE) principles

to successfully integrate predictive modelling into HILIC method development. DoE principles were first used to explore the chromatographic variables (percentage of acetonitrile, as well as pH and salt concentration) known to be effective in HILIC, followed by regression analysis to generate models capable of predicting retention parameters over a wide range of chromatographic conditions. The mathematical DoE model was shown to be highly predictive when applied to test conditions inside the design space. A QSRR model was then generated to predict retention times of test probes. A compound classification based on the concept of similarity was applied prior to QSRR modelling in order to enhance the predictive capability of QSRRs. Finally, the QSRR-DoE computed retention times of pharmaceutical test analytes and subsequently calculated separation selectivity factors were used to optimise the chromatographic conditions for efficient separation of targets. Quality assurance was achieved through the application of Monte Carlo simulation to propagate the prediction error. The desired separation for the target analytes was established experimentally, which confirmed the theoretical predictions.

In the third and main part of the thesis, an in depth study on the strategies which enhance QSRRs prediction accuracy able to support HILIC method development was carried out. A similarity searching approach was applied in order to generate localised QSRR models, in which the retention of any given compound is predicted using only the most similar compounds in the available dataset. Two similarity measures were performed; retention factor ratio as a chromatographic similarity measure and Tanimoto index as the most popular similarity measure based on chemical structure. Prediction error was reduced when QSRR was based on similar compounds rather than using the entire dataset, with an excellent result for retention time (t_R) similarity-based local models. However t_R filtering is unable to be applied to a real-life

situation, as the retention time of a new analyte is unknown. To tackle this challenge, a novel QSRR methodology was presented based on a dual-filtering strategy which combines Tanimoto similarity (TS) searching as the primary filter and tR similarity clustering as the secondary filter. To employ tR similarity filtering, correlation to a molecular descriptor was used as a measure of retention time. A comparison of diverse, global, TS-based and dual-filtering-based QSRR models over five different HILIC stationary phases showed that the proposed dual-filtering-based QSRR model was the most successful approach.

Table of content

List of abbreviations

Acronym	Representation
3DMoRSE	3D-molecule representation of structure based on electron diffraction
AC-Carn	O-acetyl-1-carnitine
ACN	Acetonitrile
AMW	Average molecular weight
AROM	Aromaticity index
ATSC1e	Centred Broto-Moreau autocorrelation of lag 1 weighted by Sanderson electronegativity
ATSC1s	Centred Broto-Moreau autocorrelation of lag 1 weighted by I-state
ATSC4s	Centred Broto-Moreau autocorrelation of lag 4 weighted by I-state
B3LYP	Becke 3-parameter (exchange) with correlation by Lee Yang and Parr
BTMA	Benzyltrimethylammonium
CA	Cluster analysis
CAD	Charged aerosol detector
CATS	Chemically Advanced Template Search
CATS2D$_{_08_DL}$	CATS2D Donor-Lipophilic at lag 08
Carn	1-carnitine
CQA	Critical quality attribute
DAD	Diode array detector
DFT	Density functional theory
DLS_02	Modified drug-like score from Oprea et al. (6 rules)
DOE	Design of experiments
ESI-MS	Electrospray ionization mass spectrometry
FA	Formic acid
FD	Fluorescence detection

Acronym	Representation
G3i	3rd component symmetry directional WHIM index / weighted by ionization potential
G3u	3rd component symmetry directional WHIM index / unweighted
G3v	3rd component symmetry directional WHIM index / weighted by van der Waals volume
GA	Genetic algorithm
GATS3e	Geary autocorrelation of lag 3 weighted by Sanderson electronegativity
GATS3p	Geary autocorrelation of lag 3 weighted by polarisability
GETAWAY	Geometry, topology, and atom-weights assembly
HATS2p	everage-weighted autocorrelation of lag 2 / weighted by polarisability
HATS3s	leverage-weighted autocorrelation of lag 3 / weighted by I-state
HATS7u	leverage-weighted autocorrelation of lag 7 / unweighted
HCA	Hierarchical cluster analysis
HILIC	Hydrophilic interaction chromatography
HOMT	HOMA (Harmonic Oscillator Model of Aromaticity index) total
HPLC	High performance liquid chromatography
Hy	Hydrophilic factor
ICH	International conference on harmonisation
IEFPCM	Integral equation formalism variant of the polarizable continuum model
J_D	Balaban-like index from topological distance matrix (Balaban distance connectivity index)
JGI2	mean topological charge index of order 2
JGI5	mean topological charge index of order 5
K	Retention factor
***K*-ratio**	Retention factor ratio
LOO	Leave-one-out
LMO	Leave-many-out

Acronym	Representation
LSER	Linear solvation energy relationship
LVs	Latent variables
MAE	Mean absolute error
MAEP	Mean absolute error prediction
MATS1e	Moran autocorrelation of lag 1 weighted by Sanderson electronegativity
MATS2i	Moran autocorrelation of lag 2 weighted by ionization potential
MATS3i	Moran autocorrelation of lag 3 weighted by ionization potential
MATS1m	Moran autocorrelation of lag 1 weighted by mass
MATS2p	Moran autocorrelation of lag 2 weighted by polarisability
MATS6p	Moran autocorrelation of lag 6 weighted by polarisability
MATS1s	Moran autocorrelation of lag 1 weighted by I-state
MATS1v	Moran autocorrelation of lag 1 weighted by van der Waals volume
MeOH	Methanol
MLOGP	Moriguchi octanol-water partition coeff. (logP)
MLR	Multiple-linear regression
MMff94	Merck molecular force field
MOPAC	Molecular orbital package
Mor22s	Signal 22 / weighted by I-state
Mor23s	Signal 23 / weighted by I-state
Mor31s	Signal 31 / weighted by I-state
Mor28u	Signal 28 / unweighted
MP	Mobile phase
MS	Mass spectrometry
NH_4Ac	Ammonium acetate
NH_4FA	Ammonium formate

Acronym	Representation
NPLC	Normal phase liquid chromatography
nVar	Number of selected variables
OVAT	One-variable-at-a-time
PCR	ratio of multiple path count over path coun
PLS	Partial least squares
PM7	Semi-empirical Parametric Method number 7
PTMA	Phenyltrimethylammonium
PW3	path/walk 3 - Randic shape index
$\mathbf{Q^2_{ext}}$	Correlation coefficient of external validation
$\mathbf{Q^2_{LOO}}$	Correlation coefficient of leave-one-out
$\mathbf{Q^2_{LMO}}$	Correlation coefficient of leave-many-out
QbD	Quality-by-design
qnmax	maximum negative charge
QSAR	Quantitative structure-activity relationship
QSPR	Quantitative structure-property relationship
QSRR	Quantitative structure-retention relationship
PCA	Principal component analysis
R5e+	R maximal autocorrelation of lag 5 / weighted by Sanderson electronegativity
R6m+	R maximal autocorrelation of lag 6 / weighted by mass
R6u+	R maximal autocorrelation of lag 6 / unweighted
RF	Random forest
RMSECV	Root mean square error of cross validation
RMSEP	Root mean square error of prediction
RP-HPLC	Reversed-phase high performance liquid chromatography
Si–H	Silicon hydride
Si–OH	Silanols

Acronym	Representation
SaasC	Sum of aasC E-states
SP	Stationary phase
SpMax_B(v)	leading eigenvalue from Burden matrix weighted by van der Waals volume
SpMax_B(m)	leading eigenvalue from Burden matrix weighted by mass
SpMax3_Bh(s)	largest eigenvalue n. 3 of Burden matrix weighted by I-state
SpPosA_B(m)	normalised spectral positive sum from Burden matrix weighted by mass
TDB08s	3D Topological distance based descriptors - lag 8 weighted by I-state
TFA	Trifuoroacetic acid
TMAO	Trimethylamine-N-oxide
t_R	Retention time
TS	Tanimoto similarity
VIP	Variable importance to projection
WHIM	weighted holistic invariant molecular

1 Introduction

1.1 Motivations and Thesis Overview

Hydrophilic interaction chromatography (HILIC) has recently become popular in the separation science community [1]. The availability of a broad range of HILIC stationary phases provides opportunities for meaningfully different retention and separation selectivity. With ever more diverse stationary phases available, it becomes more challenging for analysts to select a suitable chromatographic system or even a starting point for method development. This is especially the case in high performance liquid chromatography (HPLC) method development for the pharmaceutical industry, where the analyst has to deal with ever more complex samples containing an increasing number of individual compounds that need to be separated. One attractive solution is the development of computer-based systems which would be able to predict the retention behaviuor of the analytes with good accuracy in a particular chromatographic system. This has inspired many chromatographers to devote a great deal of effort to propose possible strategies to accelerate HILIC method development with the aid of quantitative structure-retention relationship (QSRR) methodology [2]. This thesis aims to add to this work, exploring strategies for the acceleration of HILIC method development and the computational prediction of analyte retention times.

This thesis comprises the development of retention prediction models for a variety of pharmaceutical compounds and commercially available stationary phases used in the HILIC mode. Strategies are proposed to enhance the predictive power of QSRR models using the concept of molecular similarity [3]. The proposed QSRR methodology is used in column scoping and optimising steps of HILIC method development. Furthermore, the

mechanism within HILIC columns is studied by analysis of the molecular descriptors obtained in the final QSRRs.

1.2 Hydrophilic-Interaction Chromatography (HILIC)

Reversed-phase HPLC (RP-HPLC), by far the most popular LC technique for pharmaceutical analysis, features an apolar stationary phase and a polar mobile phase. Consequently, retention increases when the polarity of the mobile phase increases and/or when the polarity of the analysed compounds and/or the stationary phase decreases [4]. The elution sequence goes from the most to the least hydrophilic (polar) compound. The main drawback of RP-HPLC is that it offers poor retention for hydrophilic compounds.

In normal phase LC (NPLC), contrary to RP-HPLC, a polar stationary phase and an apolar mobile phase are used, leading to increased retention with decreasing mobile phase polarity, and/or increasing polarity of the analysed compounds and/or stationary phase [4]. The most apolar compounds are eluted first, and the most polar last. The main drawbacks of NPLC are that the nonpolar mobile phase solvents can be quite expensive, toxic or environmentally unfriendly, and polar compounds often show poor solubility in these solvents.

The expression *hydrophilic-interaction chromatography (HILIC)* - coined by Alpert in 1990 [5] - defines an alternative chromatographic mode to RP-HPLC and NPLC. HILIC using a polar sorbent in combination with a hydro-organic mobile phase, provides an approach for the effective separation and quantitative determination of small polar compounds. Similar to NPLC, retention increases with decreased mobile phase polarity and/or with increased polarity of the analysed compounds and/or the stationary phase.

After experiencing a slow start [6, 7], in the past few years HILIC has become one of the preferred analytical techniques for the separation of polar compounds [8-11]. Recently, HILIC has been successfully applied to the analyses of a wide range of small polar compounds, including drugs, toxins, plant extracts, and other compounds important to the food and pharmaceutical industries. The reason for the increase in popularity is that HILIC offers an alternative to NP chromatography for the separation of polar compounds. HILIC mobile phases are of hydro-organic character, which permits, on the one hand, an excellent suitability for coupling to mass spectrometry (MS) detectors and especially for use with electrospray ionization mass spectrometry (ESI-MS), as acetonitrile-rich eluents assist spray formation and improve ionization efficiency, leading to enhanced detection sensitivity [12]. On the other hand, the NPLC drawback of insolubility of hydrophilic compounds is also largely solved in HILIC, because of its mobile phase properties. Moreover, HILIC has the well-known advantage of alternative selectivity to RP liquid chromatography, that is, good retention for very polar compounds (compared with low retention in RP) [13], and simultaneously, it is characterised by low operational back pressures guaranteed by the low viscosity of the organic-rich mobile phases used.

1.2.1 Retention mechanism in HILIC

Although HILIC has been widely applied, retention mechanisms of this chromatographic mode are still debated. The most accepted model is based on hydrophilic partitioning, i.e., partitioning of analytes between the mobile phase and the hydrophilic environment of the stationary phase [6, 14]. The polar surface of the HILIC stationary phases attracts water molecules, which are subsequently adsorbed to form a stagnant water-enriched liquid layer [5, 15]. The acetonitrile-rich bulk and the water-enriched layer are two liquid

phases of different polarity and can be regarded as a liquid–liquid separation system. The HILIC separation mechanism is based primarily on the differential distribution of the analyte between these two phases. The more hydrophilic the analyte, the more the partitioning equilibrium is shifted towards the immobilized water layer on the stationary phase, and thus, the more the analyte is retained [6, 16]. A schematic view of the partitioning mechanism of a hydrophilic analyte in HILIC system is provided in Figure 1.1.

It is nowadays generally accepted that retention of analytes within a HILIC system is caused by not only partition-driven phenomena but also by surface adsorption (hydrogen bonding, dipole-dipole), electrostatic interactions (attractive or repulsive) with charges on the stationary phase and also to some extent hydrophobic interactions [6, 8, 13, 14, 17-23].

1.2.2 Stationary phases for HILIC

Any polar stationary phase that can retain water may be used in the HILIC mode. With the growing interest in polar compound analysis, a large number of HILIC stationary phases have been developed on different supports (silica, polymers and hybrid materials) modified with many polar functional groups, such as amide, cyano, amino, diol, polyethylene glycol, poly(succinimide) and its derivatives, sulfoalkylbetaine, cyclodextrin, polyvinylalcohol, pentafluorophenyl-propyl, polypeptidyl and other polar functional groups [6, 24-26], which are suitable for a wide range of applications [25]. Most of the HILIC stationary phases are silica-based materials and can be arranged into roughly five groups, namely bare silica, amine, amide, diol and zwitterionic.

Unmodified silica phases remain among the most popular of HILIC stationary phases, especially in pharmaceutical analysis [6, 26-29]. A

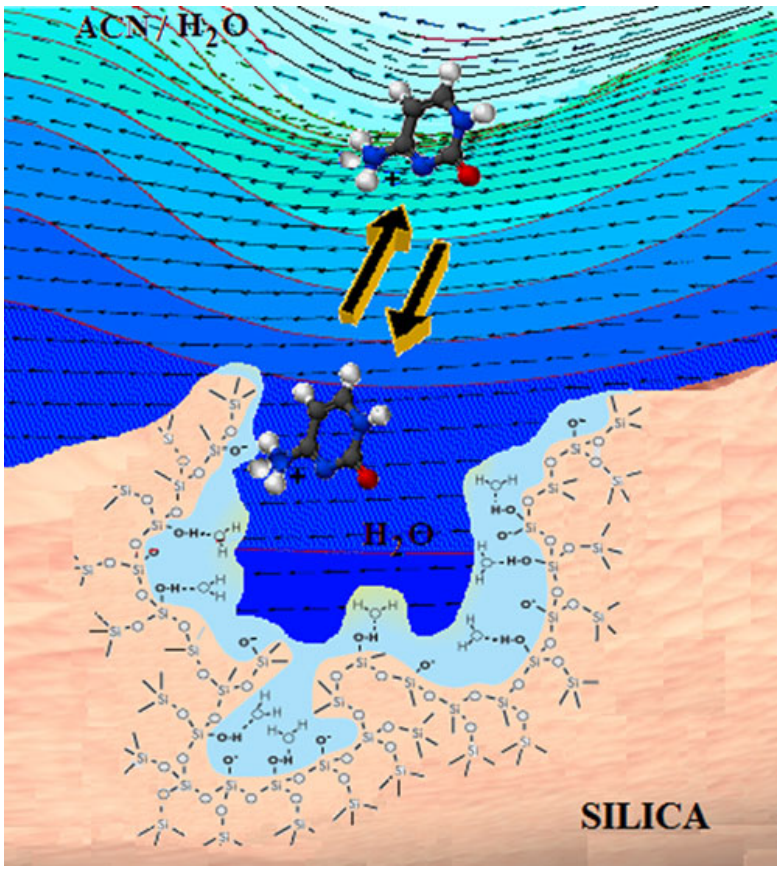

Figure 1.1. Scheme of the partitioning mechanism in a HILIC system. (Buszewski and Noga, Anal. Bioanal. Chem. 2012, © Elsevier; Reprinted with permission).

list of the applications of HILIC on unmodified silica columns in pharmaceutical analysis can be found in a recently published review [9]. Unlike chemically bonded silica phases, these phases are not subject to bleeding of the bonded phase from the column [30]. There are three types of bare silica materials, silica type A, silica type B and silica type C. Type A silica is prepared by precipitation of silicate solutions and type B silica can be prepared by the aggregation of silica sols in air [25, 31]. Unmodified bare silica gels type A and B are hydrophilic phases with silanol groups and siloxane bridges, which can display hydrogen donor and acceptor activities. At higher pH values, silanol groups are deprotonated and separation may be achieved through ion-exchange interactions for charged compounds. In this case, a cation-exchange mechanism takes place, to retain strongly positively-charged basic analytes, whereas negative analytes are poorly retained due to electrostatic repulsions. Type A and B silica materials can be contaminated with metal impurities during the preparation process. Type B silicas contain a lower amount of impurity and are more stable at intermediate and high pH compared with type A silicas. Type C silica gels are a class of less hydrophilic silica phases, consisting of a hydrosilation silica surface populated mostly with silicon hydride (Si–H) groups instead of silanols (Si–OH) groups [25]. This type of HILIC phase shows less attraction of water and consequently improved reproducibility of retention [32].

Aminopropyl-bonded silica was the first bonded stationary phase of HILIC mode separations [33] and is still widely used in the HILIC mode for separations of sugars, amino acids, peptides, carboxylic acids, nucleosides and some pharmaceuticals [25]. The primary amino group is positively charged and therefore is thought to display an ion-exchange mechanism. Uncharged and polar analytes are retained primarily through hydrophilic

interactions, whereas anion-exchange interactions prevail in the retention of charged compounds with irreversible binding for acids [34]. Amino phases with ligands containing secondary or tertiary amino groups have the advantage over aminopropyl phases in that they provide symmetrical peak shapes, shorter retention times for acids [24, 25], and longer lifetimes of the columns [25].

Chemically bonded diol phases, among the first chemically bonded silicas to be developed [35], usually have neutral hydrophilic 2,3-dihydroxypropyl ligands. The diol phase is prepared by bonding glycidoxypropyltrimethoxysilane to the silica gel surface, followed by hydrolysis of epoxy groups. Diol phases demonstrate high polarity, hydrogen bonding properties and a relative absence of ionizable groups, meaning that they are nearly ideal for the HILIC mode [36]. Hydrogen bond interactions, in addition to the hydrophilic partitioning, may play an important role in the retention of polar analytes with hydrogen donor or acceptor functionalities. Related to the diol phase is the cross-linked diol, which contains hydroxyl groups present on the surface of the polymer coating and is prepared by cross linking the diol group through an ether linkage, thus forming a polymer layer on the silica surface [37]. HILIC on a diol column is suitable for the quantitative analysis of polar active pharmaceutical ingredients in drug formulations [25].

Amide phases are produced by functionalization of the silica gel surface with carbamoyl or amide groups, linked through an alkyl spacer. Amide phases do not possess basic properties, so the retention of ionizable compounds is not affected by ion-exchange interactions. They thus show good recovery [38], repeatability and reproducibility [39]. After Yoshida [40] (producer of TSKgel Amide-80 as one of the most popular phases) applied these phases to the separation of peptides, the amide-silica phase

soon found common usage in HILIC. This type of HILIC phase is also suitable for separations of oligosaccharides, glycoproteins, or various glycosides [39].

In addition, newer stationary phases, such as zwitterionic phases, have been developed. Zwitterionic stationary phases contain equal amounts of oppositely charged groups, which impart a net surface charge equal to zero. A common zwitterionic stationary phase is the sulfoalkylbetaine phase, constituted of positively charged quaternary ammonium and negatively charged sulfonate groups separated by a short alkyl spacer. Zwitterionic ligands strongly adsorb water to the surface, and partitioning of hydrophilic analytes into the water layer largely controls the retention mechanism [41]. In addition, hydrogen bonding and electrostatic interactions may contribute to the retention of charged compounds with hydrogen donor and acceptor activities. Electrostatic interactions with the zwitterionic phases are weaker than those with the charged materials, and both anionic and cationic species can be retained by zwitterionic phases. Application examples of zwitterionic HILIC columns are the separations of small polar compounds [42], metabolomes [43], glucosinolates [44], aminoglycosides [45], glycopeptides [46] and other compounds.

Different selectivities can be obtained with amino, diol, amide and zwitterionic stationary phases, which are now becoming more popular as well as silica, and the option of different selectivities is considered to be the gold standard. The available stationary phases for HILIC have recently been reviewed by Guo and Gaiki [26] and Jandera [25].

1.2.3 Mobile phases in HILIC

As discussed above, HILIC uses mobile phases with hydro-organic character in which a high percentage of organic solvent, typically between 60-95%, is mixed with 5-40% water or aqueous buffer. HILIC retention is

strictly dependent on the composition and thickness of the water layer. Typically, a minimum of 2% of water is needed to enable the formation of the water layer on the surface of the polar stationary phase, and to create a layer which is thick enough to establish liquid-liquid partition with the bulk organic mobile phase [25]. The selection of the organic solvent has a strong effect on retention, elution order and peak shape in the HILIC mode. The ideal organic solvent should be miscible with water, but without hydrogen donor or acceptor functionalities. The elution strength of organic solvents in the HILIC mode increases by increasing solvent polarity and its ability to participate in proton-donor/proton-acceptor interactions. Relative solvent strengths in HILIC can be approximately summarised as follows: methanol > ethanol > isopropanol > tetrahydrofuran > acetonitrile (ACN) [47]. ACN is therefore the strongly preferred organic solvent in HILIC mode, while protic solvents (alcohols) that allow for hydrogen bonding interactions competing with water in the solvation of the stationary phases surface are not recommended [25, 48, 49]. Several attempts to replace acetonitrile with a less toxic solvent were reported with no success, as mobile phases containing other solvents (*e.g.* acetone, tetrahydrofuran) often provide insufficient sample retention, insufficient separation efficiency, lower intensity MS signals and significant band broadening [19, 23, 50-52]. The percentage of organic solvent in the mobile phase has a crucial influence on the retention observed in HILIC. In general, increasing concentration of the organic solvent (ACN) in the mobile phase enhances the retention of polar compounds on various types of stationary phases. The reason is that in a higher ACN concentration, a stronger interaction between the water molecules and the polar stationary phase takes place, allowing a stronger partitioning mechanism [25].

As mentioned above, the ion-exchange interaction on the charged stationary phases contributes significantly to the retention and selectivity in

HILIC. Early research indicated that the presence of buffer salts in the mobile phase allowed control of electrostatic interactions [53]. Salts commonly used in HILIC are ammonium acetate and ammonium formate due to their good solubility in mobile phases with high organic content and compatibility with mass spectrometric detection. The effect of the type and concentration of buffer salts on the retention of polar compounds on various polar phases has been investigated in several HILIC studies [8, 19, 20, 27]. In general, the presence of buffer salts in the mobile phase can effectively reduce the electrostatic interactions (both attractive and repulsive) of charged solutes with charged or zwitterionic HILIC stationary phases. In the case of electrostatic attractions, an increase in the salt concentration leads to decreased retention of charged solutes on stationary phases of opposite charge, whereas in the case of electrostatic repulsions, it leads to increased retention of charged solutes on stationary phases with the same charge [8].

Mobile phase pH is an important chromatographic factor since it can affect the charge state of both the stationary phase and the polar solutes. HILIC separations are usually performed with mobile phases in the pH range of 3–8, and formic and acetic acids are the common acid additives due to their volatility and MS compatibility. The residual silanols on the surface of silica-based stationary phases are significantly deprotonated at pH above 5, leading to an increase in electrostatic interactions between the negatively charged stationary phases and positively charged solutes [54].

HILIC separations are performed either in the isocratic mode with a high percentage of acetonitrile or with gradients starting from 95% acetonitrile containing 5% aqueous ammonium acetate or ammonium formate buffer and ending with a high water concentration (up to 90%) to remove strongly retained sample compounds [16, 25].

1.2.4 Application of HILIC in pharmaceutical analysis

A wide variety of pharmaceutical compounds, including therapeutic, diagnostic, and preclinical drugs, as well as toxic compounds, have been analysed under the HILIC mode [9]. Pharmaceutical compounds and their metabolites often have ionizable functional groups, resulting in insufficient resolution, long equilibration requirements, poor retention and co-elution when applying the RPLC mode for separation [55]. HILIC is considered suitable to analyse the polar drugs and their metabolites, with higher signal-to-noise ratios and thus better sensitivity, better retention and better peak shapes [55]. The application of HILIC analysis for some pharmaceutical compounds from publications in 2016 is summarised in Table 1.1.

1.3 Retention prediction in HILIC

Retention prediction modelling plays an important role in chromatographic method development. Several approaches are generally employed in retention modelling in HILIC, such as retention-mechanism-based, quantitative structure–retention relationship (QSRR)-based, linear solvation energy relationship (LSER)-based and hydrophilic-subtraction-based modelling approaches.

1.3.1 Mechanism-based models

In HILIC, the mechanism-based approach has mostly been used to discuss how retention behaviour is affected by the mobile phase composition, however the quantitative description of retention has received little attention [6, 41]. Many studies on the retention mechanism have centred around two common models, namely, partitioning and adsorption as described by eq (1.1) and eq (1.2) [74, 75]:

$$\log k = logkw - S\varphi \qquad (1.1)$$

Table 1.1. Conditions for analyses of pharmaceutical analytes under HILIC mode

Analytes	Column		Mobile phase	Detection	Ref.
	Column name	Functional group			
phenolic acids	Unitary XAmide	amide	ACN/water/FA[1]	Photo DAD[2]	[56]
cotinine	LunaHILIC	diol	ACN/NH_4Ac[3] buffer (pH 5.8)	DAD	[57]
neuro-transmitters	XBridge Amide™ BEH	amide	ACN/NH_4FA[4] buffer (pH 3.0)	MS	[58]
lipids	Cogent Diamond Hydride	silica	ACN/FA/NH_4FA buffer (pH 4.0)	MS	[59]
amino acids	Accucore Amide HILIC	amide	ACN/NH_4FA buffer (pH 3.2)	CAD[5]	[60]
insulins	ACQUITY UPLC Glyco-protein BEH Amide	amide	ACN/TFA[6]/water	UV-DAD	[61]
veterinary drugs	ACQUITY UPLC BEH HILIC	hybrid silica	ACN/MeOH[7]/NH_4FA	MS/MS	[62]
trisulfide	Acclaim Trinity P1 column	hybrid silica	ACN/NH_4FA buffer (pH 4.0)	CAD	[63]
β-blockers	ZORBAX RRHD Poroshell 120-HILIC	silica	ACN/NH_4Ac buffer (pH 7.5)	UV	[64]
biothiols	ZIC-HILIC Inertsil Amide	sulfoalkyl-betaine amide	ACN/NH_4FA buffer (pH 3.0)	FD[8]	[65]
choline, betaine, AC-	Phenome-nex Luna HILIC	diol	ACN/water/NH_4Ac buffer (pH 4.0) and	MS/MS	[66]

Carn[9], Carn[10] and TMAO[11]			ACN/ NH_4FA buffer (pH 3.0)		
cetirizine	Poroshell 120 HILIC	silica	ACN/FA	MS	[67]
miltefosine	Phenomenex Luna HILIC	diol	MeOH/FA	MS/MS	[68]
nicotine	Phenomenex Luna HILIC	diol	ACN/NH_4FA buffer (pH 3.2)	MS/MS	[69]
granisetron	ZIC-HILIC	sulfoalkyl-betaine	ACN/NH_4Ac buffer (pH 3.0)	UV	[70]
glycans	XBridge BEH Amide	amide	ACN/NH_4FA	MS	[71]
genotoxic impurities	Kinetex HILIC, Phenomenex XBridge HILIC, and Primesep B	silica, silica, and amine	ACN/NH_4FA/FA	Photo DAD	[72]
amino acids	ACQUITY UPLC BEH amide	amide	ACN/NH_4FA/FA	MS	[73]

[1]formic acid, [2]diode array detector, [3]ammonium acetate, [4]ammonium formate, [5]charged aerosol detector, [6]trifluoroacetic acid, [7]methanol, [8]fluorescence detection, [9]O-acetyl-l-carnitine, [10]l-carnitine, [11]trimethylamine-N-oxide.

$$logk = logk_B - \frac{A_S}{n_B} logC_B \qquad (1.2)$$

where φ is the volume fraction and C_B the mole fraction of solvent B (stronger solvent) in the mobile phase, k_w is the extrapolated value of retention factor when φ = 0, and k_B is the retention factor when C_B = 1. S is the slope of log k *versus* φ when fitted to a linear regression model. A_S and n_B are the cross-sectional areas occupied by the solute molecule on the surface and the B molecules, respectively.

The mechanism-based approach has also been used in modelling the retention of polar solutes in HILIC. Eq (1.1) represents the simplest partitioning model based on linear solvent strength theory, but it is only valid for narrow φ ranges. It relies on the assumption that there is a linear relationship between $logk$ and the mobile phase composition. Two modified partitioning models, a quadratic model (eq (1.3)) [76] and an empirical model (eq (1.4)) [77], have been proposed to describe retention behaviour more accurately than the linear model (eq (1.1)):

$$\ln k = lnkw + S_1\varphi + S_2\varphi^2 \qquad (1.3)$$

$$\ln k = lnkw + 2\ln(1 + S_2\varphi) - \frac{S_1\varphi}{1 + S_2\varphi} \qquad (1.4)$$

In addition, Liang and coworkers proposed a mixed model (eq (1.5)) to describe the retention behaviour of polar compounds in HILIC [78]:

$$\ln k = lnkw + S_1\varphi + S_2 ln\varphi \qquad (1.5)$$

where φ is the fraction of water. They have applied the proposed quantitative descriptive retention equation to predict retention time of studied compounds under different mobile phase conditions. They investigated the retention of 8 nucleosides on 6 polar stationary phases with different functional groups and compared four retention models (eq (1.1), eq

(1.2), eq (1.3) and eq (1.5)). The mixed model seemed to fit the retention data better than the partitioning and adsorption models. Guillarme and co-workers [79] investigated the possibility of retention modelling in the HILIC mode, testing equations (1.3), (1.4) and (1.5) using data generated with 9 compounds on 4 different HILIC stationary phases (a bare silica, a hybrid silica, an amide and a zwitterionic column). Among the considered mathematical models, the mixed model (eq (1.5)) was found to best fit the retention data and the quadratic model (eq (1.3)) gave the poorest fit in the isocratic mode, while the empirical model (eq (1.4)) seemed to be the best compromise for HILIC gradient mode prediction.

1.3.2 Quantitative structure-retention relationships (QSRR)

The fundamental assumption in QSRRs is that a relationship exists between molecular descriptors and retention parameters [80]. QSRR studies start from the selection and calculation of descriptors characterising the molecular structure of a series of analytes, followed by the use of computerized statistical chemometric techniques to derive mathematical models of retention parameters as a function of the molecular descriptors. Details of QSRR methodology are discussed in Section 1-5. QSRRs have been utilised to good effect: Jinno et al. [81] reviewed the application of QSRR methods in predicting the retention factors of adrenoreceptor agonists and antagonists in HILIC using predefined solute and mobile phase descriptors. Kaliszan et al. [82] developed an accurate QSRR model to predict the retention times of compounds (metabolites and drugs) analysed in the HILIC mode using multiple-linear regression (MLR). The obtained model allowed false positive identification to be removed during the interpretation of metabolomics data. Creek et al. [83] established a QSRR model incorporating six predefined physicochemical variables in a MLR model based on 120 authentic standard metabolites with good predictive

ability. Application of the model led to an improvement in metabolite identification. More recently, Cao et al. modelled the retention time in HILIC using a random forest (RF) algorithm for the purpose of peak annotation of plant metabolites [84].

1.3.3 Linear solvation energy relationships (LSER)

The LSER is known as a QSRR model that correlates the retention parameter of solutes to their characteristics as described by the Abraham parameters [85]. The conventional representation of the LSER model presented by Abraham et al. [85] is given in eq (1.6).

$$logk = c + eE + sS + aA + bB + vV \quad (1.6)$$

in this equation, capital letters represent the solute descriptors, related to particular interaction properties, while lower case letters represent the coefficients of the model or system constants, related to the complementary effect of the stationary phases. The model intercept term is c, which when the retention factor is used as the dependent variable is dominated by the phase ratio. The terms E, S, A, B, and V are solute-dependent molecular descriptors. E is the excess molar refraction and expresses polarisability contributions from n and π electrons; S is the solute dipolarity and polarisability; A and B are the solute overall hydrogen-bonding acidity and basicity; and V is the McGowan characteristic volume. To be able to apply the LSER model in the HILIC mode with a mixed mode mechanism, Chirita et al. [20] modified the conventional LSER model with two additional parameters which account for the electrostatic interactions of the charged solutes, as shown in eq (1.7)

$$logk = c + eE + sS + aA + bB + vV + d^{-}D^{-} + d^{+}D^{+} \quad (1.7)$$

where D^{-} represents the negative charge carried by anionic solutes and D^{+} represents the positive charge carried by cationic solutes. The LSER model

has been successfully used for RPLC [86], however the statistical analysis of the regression equations showed that LSER is not accurate enough to support the retention prediction purpose in HILIC and its application is limited to column classification [20, 87].

1.3.4 Hydrophilic-subtraction model

The research group of Snyder and Dolan has published a series of articles dealing with the characterisation of a variety of RP stationary phases using a hydrophobic-subtraction model [88]. The column parameters have been determined for a wide range of columns, and are available from the U.S. pharmacopeia [89]. This group classified and compared stationary phases, providing a means to select stationary phases of different characteristics for method development. Other research groups [90] have also classified columns using the hydrophobic subtraction method in RP mode.

More recently, Liang and co-workers developed a hydrophilic-subtraction model to describe the retention in HILIC, which is constructed using the same strategies applied to produce the hydrophobic-subtraction model [91]. The hydrophilic-subtraction model was designed based on the major interactions governing HILIC retention, including hydrophilic partitioning, hydrogen-bonding and electrostatic interactions as shown in eq (1.8):

$$logk = logk_{ref} + hH + aA + bB + cC + dD \qquad (1.8)$$

where k is the retention factor of a given solute, and k_{ref} the value of k for a reference compound on the same column under the same conditions. In eq (1.8), all the capital letters are solute descriptors: H, hydrophilicity; A, solute hydrogen-bond acidity; B, solute hydrogen-bond basicity; C, solute cation-exchange activity; and D, solute anion-exchange activity. The lower-case letters (h, a, b, c, d) are the system constants reflecting the magnitude

of difference in the particular interactions between the stationary phase and the mobile phase. A multiple linear regression was performed to calculate all the solute descriptors and system constants. This approach was successfully applied for HILIC column classification purposes. The high correlation coefficients ($R^2 \geq 0.990$) of the hydrophilic-subtraction model may indicate the reliability of this model for retention prediction of analytes.

1.4 HILIC method development

1.4.1 Chromatography method development

In the development of a chromatographic method three stages can be considered. The first stage is to select the appropriate chromatographic technique, which provides the desired separation selectivity for the particular analytes to be separated *(method selection)*. The next stage (stage 2) is to choose the particular combination of SP (i.e. which chromatographic column), MP (i.e. which liquid or mixture of liquids), and type of flow-through detector, which is most likely to lead to a successful separation. In this step, the retention of the substances is often optimised by selecting a mobile phase with an acceptable solvent strength *(retention optimisation)*. Retention has to be sufficiently low to obtain an acceptable analysis time, but also sufficiently high to achieve separation of the compounds of interest. Conditions leading to an acceptable retention do not necessarily lead to the separation of all peaks.

The final stage (stage 3) of method development is to identify the precise details of the separation, such as the exact MP composition, the flow-rate, the column length, the temperature, etc. In this stage, first, organic modifier composition is adapted (by *e.g.* replacement of one organic modifier by another one, or a change in the pH and/or the ionic strength of the buffer in the mobile phase) in order to achieve selectivity *(selectivity optimisation)* followed by the optimisation of the system *(system optimisation)*. In the

latter step, system parameters such as the column length, the particle size and the flow-rate can be changed in order to further improve the resolution or the sensitivity of the method, or to reduce the analysis time while a similar separation is maintained. In the system optimisation phase the best experimental conditions for a sufficient resolution of the relevant peaks that gives acceptable and preferably robust results in a short analysis time are defined. For clarity, throughout this thesis we will refer collectively to stages 1 and 2 as *"scoping"* the LC method, and stage 3 as *"optimising"* the LC method.

The traditional approach (one-variable-at-a-time (OVAT) approach) to LC method development is to systematically vary one parameter at a time while keeping the others constant. The best performing level of the varied parameter is identified normally by visual inspection of the trial chromatograms; the parameter is then fixed at this level, and a new parameter is selected for the next iteration. In the OVAT process, factors are examined sequentially until an adequately performing instrumental method is obtained. At present, most of the method development undertaken in the pharmaceutical and other industries is carried out by the OVAT process, which is not only time-consuming and costly but can lead to chromatographic methods that are not inherently robust and are poorly understood. That is, a small change in an experimental parameter can lead to a major loss in performance of the method. This OVAT process requires screening of multiple types of LC techniques and numerous chromatographic columns using a large number of experimental conditions, which generates significant waste of resources (both human and instrumentation) as well as excessive consumption of organic solvents. New advances in combinatorial chemistry have enabled pharmaceutical industries to synthesize a much greater number of potential new drugs than at any time

in the past. As a result, analytical method development is also required to be fast enough to keep up with the high-throughput drug discovery processes.

1.4.2 Column scoping

Hundreds of HILIC stationary phases have been tested and characterised in the literature. Several approaches for the analysis of the stationary-phase chemistry have been established, and can be divided into chromatographic test methods and chemometric methods. The chromatographic tests can be further divided into the determination of physico-chemical properties and model-based tests. The information obtained from column classification helps prospective users to select columns needed for their specific purposes.

1.4.2.1 Determination of physico-chemical properties of the stationary phase

This method of column characterisation is helpful to classify HILIC stationary phases on the basis of their chromatographic properties. Such properties can be determined using retention/separation information of given solutes that are known to reflect given stationary phase properties. Kawachi et al. [17] have analysed 14 commercially available HILIC columns in terms of their degree of hydrophilicity, selectivity for hydrophilic/hydrophobic substituents, regio-selectivity and configurational differences in hydrophilic substituents, selectivity of molecular shapes, evaluation of electrostatic interactions and evaluation of the acidic/basic nature of the stationary phase using model nucleoside, phenyl glucoside and xanthine derivatives as probe analytes. They used radar plots to classify the stationary phases and to illustrate the characteristics of each phase visually. Ibrahim et al. [92] have characterised 30 HILIC stationary phases by constructing simple selectivity plots capable of classifying different HILIC phases based on their hydrophilicity and ion-exchange properties.

1.4.2.2 Model-based column characterisation

This characterisation is based on building a specific model, such as the linear solvation energy relationships or the hydrophobic subtraction model. The model coefficients provide information on the properties of the column, such as *e.g.* its hydrophilicity, hydrogen-bonding acidity and basicity, electrostatic interactions and steric resistance to the insertion of bulky molecules into the stationary phase. Schuster et al. [87] employed the LSER [20] approach to characterise 22 commercially available and home-made HILIC columns. Hierarchical cluster analysis (HCA) according to the obtained normalised LSER coefficients revealed that the stationary phases could be arranged into three main groups: acidic, basic and neutral. More recently, Wang et al. [91] have established a hydrophilic-subtraction model using a set of 41 solutes and 8 HILIC columns. A hydrophilic-subtraction model correlates the retention with solute descriptors and column parameters. The model was successfully validated by characterising 15 test HILIC columns. The model regression coefficients were used to characterise HILIC stationary phases by an angle graph and a spider diagram, which could be used as guidance for researchers to select appropriate columns for their specific purposes.

1.4.2.3 Chemometric methods

Chemometric methods can be used to handle and interpret data sets from chromatographic tests and do not directly analyse stationary phase chemistry or properties. Such methods include principal component analysis (PCA) and cluster analysis (CA), which help to classify columns in groups based on similar (or dissimilar) properties. This can then simplify column selection procedures. Chirita et al. [54] have considered several neurotransmitters as model compounds and they have evaluated the retention data of 12 HILIC columns by performing PCA calculations. They

have then proposed a method development scheme useful to help in the initial choice of a HILIC stationary phase for the separation of polar targets. Lammerhofer et al. [93] have evaluated 19 mixed-mode columns based on their hydrophobic selectivity and ion-exchange capacity under HILIC and RPLC modes using PCA calculations. They have employed xanthines, nucleosides and vitamins as model analytes. Dinh et al. [18] conducted an extensive study on 22 hydrophilic and polar HILIC stationary phases using PCA calculations, taking into account their hydrophilic, hydrophobic, electrostatic, hydrogen bonding, dipole–dipole, π-π interaction and shape-selectivity interactions, using a set of 21 hydrophilic probe compounds. They have classified the columns into four functional groups based on their selectivity: (1) cation-exchange – unmodified silicas; (2) anion-exchange – columns with amino functionality; (3) dipole-dipole and multi-point hydrogen bonding – polymeric sulfobetaine and polysulfoethyl phases; (4) low specific interaction group – hydroxyl, diol, amide, and monomeric zwitterionic phases. More recently, Periat et al. [23] have provided some guidelines for HILIC method development in pharmaceutical analysis. For this purpose, they have analysed five HILIC stationary phases (bare silica, hybrid silica, amide, diol and zwitterionic) under various mobile phase conditions based on PCA, HCA and correlation maps, employing a diverse set of 82 pharmaceutical compounds. The chemometric analysis revealed that the diol and hybrid silica phases were similar in both retention and selectivity, and the bare silica and amide phases appeared to be in separate groups in terms of selectivity, while the zwitterionic phase was found to behave differently from the other phases.

1.4.3 Method optimisation

Method optimisation can, generally, be conducted in two ways, univariate or multivariate. As mentioned above, usually the OVAT approach

is used, which is univariate in nature. Various publications on univariate HILIC method development reported in the literature have been reviewed by Dejaegher et al. [94]. A drawback of these methods is that they are time-consuming due to the large number of experiments that must be conducted in order to determine the effect of each parameter on the retention of the analytes. Thus, the need to maximise the efficiency of scientific discovery in order to minimize waste and cost, has caused researchers to do smarter experiments that give the most information possible with the fewest experiments.

An alternative to the single variable approach is the use of a Quality-by-Design (QbD) methodology for LC method development [95, 96]. A QbD methodology for LC method development uses a statistical experiment design plan (Design of Experiments, or DOE) [97] to systematically vary multiple study factors in combination through a series of experiment trials that, taken together, can comprehensively explore a multi-factor design space. Such a design can provide a data set which can be used to identify and quantify interaction effects of the factors. This quantitation translates the design space into a knowledge space. This increased depth of knowledge leads to the possibility that the process can be designed to operate robustly and to be capable of tolerating small changes in experimental variables without loss of performance. In turn, this allows the developed process to be transferred successfully between locations where both personnel and instrumentation differ. When applied to chromatographic methods, QbD principles dictate that the specific experimental conditions used for a particular separation should be able to provide and maintain optimal performance, even when there are slight changes in these conditions [98]. A general QbD guide for separation method development is shown in Figure 1.2.

While the classic theories of experimental design have been around since the middle of the twentieth century, adoption of DOE methods in chromatography research has seen increased activity only in the past decade. Although experimental design-based procedures have already been applied successfully to optimise other analytical methods, *e.g.* RP-HPLC and NPLC methods, unfortunately, only a few HILIC methods found in the literature have been optimised using an experimental design approach [23, 26, 99-102]. The DOE method is a planned series of experiments with changing variables describing the experiment in the most efficient way [103]. The aim of DOE is to get the best description of the response surface, which is a 2D or 3D plot showing the influence of one or more variables on an output response.

In an experimental design approach, different steps can be distinguished. First, the factors to be examined should be defined as well as the level intervals in which these factors will be evaluated. Factors to be optimised in the HILIC mode using the DOE method can include the acetonitrile content in the mobile phase, buffer concentration and pH of the mobile phase, and column temperature [99, 100]. After choosing factors, a screening design is usually applied that allows the examination of a relatively high number of factors in a feasible number of experiments [104]. The factors most influencing the assay can then be determined by analysing the results of the screening design. Subsequently, these factors then can be further evaluated using a response surface design methodology [104], in which a relationship between responses to values of one or more factors is established. In the response surface design step, an optimisation design is selected which is determined by the definition of the responses. For separations, responses related to the quality of the separation/method, such as resolution and retention factor, are usually selected. In a further optimisation phase, the experimental protocol is defined and the required experiments are

performed. For each experiment, the responses are measured or calculated. Consequently, data analysis is performed on the design results. For a response surface design, the response is usually modelled by a second order polynomial model and then a 2-D contour plot or a 3-D response surface can be drawn to visualize and to determine the best and most robust conditions.

Before an optimised method can be used in routine analysis, *method validation* is performed to show that the method is capable of doing what is claimed. Before starting with method validation often a robustness or ruggedness test is performed. In a *robustness/ruggedness test* one evaluates the influence of small variations in the procedure on the performance of the method. These small variations are deliberately introduced and represent variations that could occur when a method is transferred, *e.g.* from one laboratory to another.

1.5 Quantitative structure-retention relationships (QSRRs)

QSRR methods represent a powerful tool in chromatography and are intensively studied for chromatographic retention predictions [80, 105, 106]. QSRRs, as the name suggests, are techniques for relating the variations in one response variable (***Y***-variable) to the variations of several descriptors (***X***-variables), with predictive purposes. ***Y***-variables are often called dependent and ***X***-variables independent variables. One of the ***Y***- or ***X***-variables should be related to chromatographic retention, the other should encode the molecular structure. QSRR methodology assumes that chromatographic behaviour of analytes is correlated with the chemical structure and that as a consequence chromatographic characteristics can be modelled as a function of calculable molecular descriptors. QSRR methodology is used for the solution of the following problems in analytical chemistry: (i) the identification of the most informative structural descriptors, (ii) understanding of molecular mechanisms of separation under

chromatographic conditions, (iii) the quantitative comparison of separation properties of chromatographic columns, (iv) the prediction of retention of new compounds or identification of unknown compounds.

A scheme of the QSRR methodology is shown in Figure 1.3. A typical QSRR study comprises the following steps: compilation of a database of retention of analytes of known structures, structure entry in order to estimate descriptors, descriptor selection, model building, and model validation.

1.5.1 QSRR components

1.5.1.1 Molecular descriptors

There are three common ways of structure representations, whole molecule 1D descriptors (sometimes known as 0D), 2D descriptors, and 3D descriptors. 1D descriptors attempt to express chemical information in a simple 1D molecular code and are designed for compact storage of information. These pre-defined codes allow the measurement of whole molecule descriptors, which represent the bulk molecular properties. 2D descriptors are computed from a chemical structure that is represented as a connection table or a molecular graph. In the graphical representation of molecular structures, atoms in the molecular structure are represented as vertices while bonds are represented as edges. 3D molecular descriptors provide molecular information about the 3D arrangement of structural features and general molecular surfaces and volumes. There are many thousands of descriptors defined in a comprehensive handbook [107].

Dragon software [108] is widely used to calculate molecular descriptors for QSRR modelling. Generally, the Dragon calculated descriptors encoding the molecular structure are classified in 22 different types: constitutional (1D), functional group counts (1D), atom-centred fragments (1D), charge

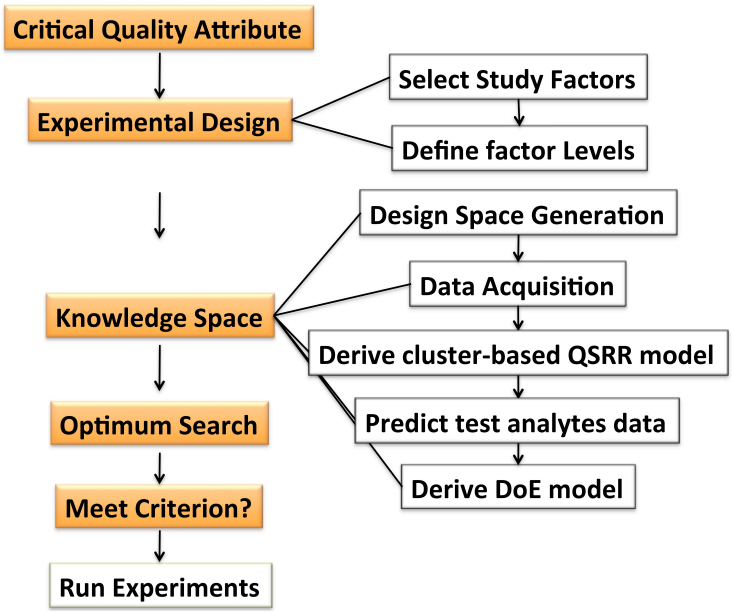

Figure 1.2. Quality-by-design guide for analytical method development.

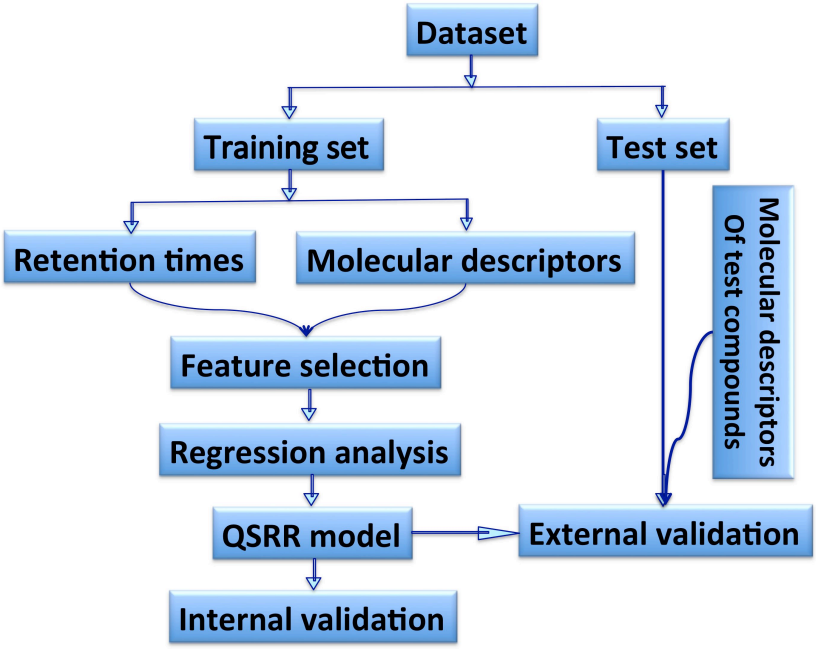

Figure 1.3. scheme of the QSRR methodology.

descriptors (1D), molecular properties (1D), topological (2D), walk and path counts (2D), connectivity indices (2D), information indices (2D), 2D autocorrelation (2D), edge adjacency indices (2D), Burden eigenvalues (2D), topological charge indices (2D), eigenvalue-based indices (2D), 2D binary fingerprints (2D), 2D frequency fingerprints (2D), Randic molecular profiles (3D), geometrical descriptors (3D), Radial Distribution Function (RDF) descriptors (3D), 3D-MoRSE (3D Molecular Representation of Structure based on Electron diffraction) descriptors (3D), WHIM (Weighted Holistic Invariant Molecular) descriptors (3D), and GETAWAY (Geometry, Topology and Atoms-Weighted AssemblY) descriptors (3D).

Chemometric methods are utilised to identify the subset of descriptors that shows the strongest ability to predict retention times (feature selection) and to derive a mathematical relationship between analyte descriptors and retention time (regression technique). The goal is to use the smallest number of analyte descriptors commensurate with a valid prediction of retention time (see below).

1.5.1.2 Feature selection

One of the most important processes in QSRRs is feature selection, which is applied to select the most relevant variables (descriptors) from a large pool. Feature selection is important because some variables in a given data set may be redundant, be irrelevant or represent noise [109, 110]. Feature selection can provide faster and more cost-effective models by helping to avoid overfitting when modelling, improving the model performance, and reducing the model dimensions [109, 110]. In addition, feature selection can help in interpretation of the retention mechanism by gaining knowledge about the most important features [72]. Many feature-selection techniques have been applied in QSRR studies, such as genetic algorithms on MLR (GA-MLR) [111], partial least squares regression [112,

113], artificial neural networks [114, 115], support vector machines [116], classification and regression trees [117, 118], and random forests [118].

1.5.1.3 Regression analysis

Multiple linear regression (MLR) was the first statistical tool used widely for QSRR models [105, 106]. With the introduction of additional molecular descriptors, many new chemometric modelling tools have been applied to QSRR modelling in order to handle the greater number of variables [119].

Partial least squares (PLS) regression is a linear, multiple regression method used frequently in chemometrics and multivariate calibration studies. Unlike MLR, PLS is particularly useful in reducing the dimensionality of the large set of independent molecular variables even in the presence of co-linearity, redundancy, and noise in both independent and dependent variables [120, 121]. In mathematical terminology, PLS summarises the variation in the independent variables into a small set of orthogonal, linear, latent variables (LVs) by maximising the covariance between molecular descriptors and the dependent variable [122, 123]. The complexity of the model is controlled by optimising the number of LVs, thus over-fitting can be minimized. The PLS method can be expressed as

$$y = a_1LV_1 + a_2LV_2 + \cdots + amLV_m \qquad (1.9)$$

where y is the dependent variable, $a_1, a_2, \ldots, a_m$ are the regression coefficients, and LVi is the ith latent variable, that is, a linear combination of the independent variables x_i

$$LV_i = b_1x_1 + b_2x_2 + \cdots + b_nx_n \qquad (1.10)$$

where x_i is the ith independent variable and bi is the ith variable coefficient.

1.5.1.4 Model validation

The statistical reliability of QSRR models should be validated, and this can be performed by several approaches including leave-one-out cross validation, leave-many-out cross validation, y-randomization, and external validation.

Leave-one-out (LOO) cross validation [119] is one of the simplest procedures for model validation. It consists of excluding each analyte once, constructing a new model without this compound, and then predicting the value of its retention parameter. Therefore, for a training set of M samples, LOO is carried out M times, resulting in M predicted values. The residuals - the differences between the experimental and predicted values from the model - are used to calculate the root mean square error of cross validation (RMSECV) and the correlation coefficient of leave-one-out cross validation (Q^2_{LOO}), as presented in eq (1.11) and eq (1.12), respectively. The minimum acceptable statistics for regression models in QSRR include conditions $Q^2 > 0.5$ [124].

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{n}(yi(obsd) - yi(pred))^2}{n}} \qquad (1.11)$$

$$Q^2 = 1 - \frac{\sum_{i=1}^{n}\left(yi(obsd) - yi(pred)\right)^2}{\sum_{i=1}^{n}\left(yi(obsd) - y(mean)\right)^2} \qquad (1.12)$$

where yi(obsd) and yi(pred) are the observed and predicted retention times of the left out analyte from the training set during cross validation. The y(mean) is the average value of the observed retention times and n is the number of analytes.

Leave-many-out (LMO) cross validation [125, 126] is performed by partitioning the data set randomly into a validation data set consisting of d analytes and the construction data subset, which contains the remaining (M

– d) compounds. The data splitting process is repeated many times and the cross-validated error estimates are averaged over all data splits. LMO cross-validated correlation coefficients (Q^2_{LMO}) are calculated in the same way as for LOO (eq (1.10)). It is recommended that d represents 20-30% of the dataset [127], and it has been shown [128] that repeating the LMO cross validation for randomly ordered data and using average of Q^2_{LMO} is statistically more reliable than LMO being performed only once.

The purpose of the y-randomization test [129] is to detect and quantify chance correlations between the dependent variable and the descriptors. The y-randomization test consists of several runs for which dependent variable values are scrambled and new QSRR models are developed using the original descriptors values (unrandomized or real values). For an acceptable QSRR model, the predictive ability of the real model should be stronger than that of the randomized models.

External validation testing requires the split of the complete data set into training and external validation sets. The purpose of this validation is to test the predictability of the QSRR model. Basic statistical parameters that are used to judge the external validation performance are the root mean square error of prediction (RMSEP), and the correlation coefficient of external validation (Q^2_{ext}), which are calculated in the same way as for internal validation, eq (1.11) and eq (1.12) respectively. Here $y_{i(obsd)}$ and $y_{i(pred)}$ are the observed and predicted retention times of the test compounds.

1.5.2 QSRR accuracy

The predictive accuracy of a QSRR model is influenced by the following basic operations: (i) the feature-selection method applied to select the most relevant descriptors, (ii) the modelling approach used for the building of the QSRR, (iii) the model validation approach (*e.g.* splitting the dataset into

training and test set), (iv) the number of predictor variables (descriptors) that are incorporated into the model, (v) the geometry optimisation method, (vi) the size of whole dataset, and (vii) the range of diversity or similarity of the molecular structures or characteristics.

One concern occurs when a diverse data set, covering a wide range of chemicals, is utilised to find a global QSRR model. The retention time prediction accuracy may not be sufficient to provide a reliable method because such a diverse dataset with different molecular characteristics and structures is forced into a general model. QSRR models based on compound-classification may provide greater predictive power compared with the QSRR models derived from the diverse dataset. This compound-classification based QSRR modelling strategy has been validated in a few papers. For example, Wang et al. [130] presented subset-specific models by using a compound classification method based on log D profile similarity. The result showed that while the predictive accuracy for QSRR models derived from subsets consisting of molecules with similar log D profiles might not have improved compared to traditional QSRR models, there was improved elution order prediction in acidic and basic chromatographic conditions. Muteki et al. [131] have also assessed the reliability of the QSRR prediction and found that the compound-classification based QSRR methodology significantly improved the retention time predictability of compounds by comparing with the global models.

The present thesis presents a novel scheme of compound-clustering based QSRR modelling based on the concept of molecular similarity.

1.6 The concept of molecular similarity

The essence of the molecular similarity concept, which originates from the comprehensive book “In concepts and applications of molecular similarity” written by Johnson and Maggiora [132], is that structurally

similar compounds are more likely to exhibit similar properties. From this, an area of interest has been the prediction of properties of chemical compounds on the basis of molecular similarity [133-137]. The term similarity is used widely for determination of quantitative relationships between the structure and properties of compounds (QSPR analysis) through cluster analysis [134, 137], similarity searching [136] and diversity analysis [138]. Cluster analysis and similarity searching allow the use of a similar subset as a training set as an alternative to the diverse training set traditionally used in QSRR studies. While comparing diverse and similar datasets for the modelling of structure-retention relationships, one may notice that in a diverse dataset method, the model reflects the properties of diverse compounds; therefore, a novel compound may be dissimilar to the compounds in the training set and the target may have properties that are not found in the training set. On the other hand, in the case of a similar dataset, the model takes into account the properties of similar compounds to the novel compound.

The degree of resemblance between two compounds is calculated using a similarity measure, which has three components: (i) a structural representation, used to represent the two compounds that are to be compared, (ii) a weighting (standardisation) scheme, used to normalise the different parts of the chosen representation, and (iii) a similarity coefficient, used to calculate the degree of resemblance between the molecules' representations. There is a comprehensive review on the three components of similarity by Willett [3] and this thesis hence briefly discusses these three components.

1.6.1 Chemical representations

In similarity calculations, a combination of 1D molecular descriptors may be used to represent a relevant property, normally after some sort of

standardisation procedure. An overview of 1D similarity measures is given by Dixon [139].

2D molecular descriptors are a straightforward chemical representation used to easily calculate the structural similarity between molecules. Most approaches using 2D structural description to quantify chemical similarity can be grouped into two broad classes. First, there are topological indices, which are single numbers that typically characterise a structure based on its size and shape [140]. The most well-known topological indices and their applications in chemistry are described in the reviews by Balaban and Estrada [141, 142]. Second, there are some approaches in which a molecule is described based on some count of shared substructure features, either by the molecule's connection table [143-146] or the fragment substructure encoded by setting-bits in a bit-string (or fingerprint) [147, 148].

Different types of fragment coding systems are available, which fall into two major classes: dictionary based fingerprints and open-ended list based fingerprints. The dictionary-based fingerprints represent dictionaries of predefined structural fragments. Structural 'keys' [149] is the simplest form of substructure fingerprints; each bit position in the string is associated with a specific molecular feature and is, therefore, dependent on a predefined list of structural fragments. In the dictionary-based fingerprint, the presence of individual fragments in a structure can be described as a sequence of 0s and 1s, where 0 means that the fragment is not present in the structure and 1 means that it is present.

The open-ended list based fingerprints involve the hashing of unique structural paths. In this approach, individual bit positions are characterised by the nature of the chemical structures in a database rather than specific features in some database. Daylight Chemical Information [150] and ChemAxon hashed fingerprint [151] are typical examples of a hashed

fingerprint encoding method, which encode molecular features into a bit string of fixed format by applying a hash function to map possible paths through the molecular graph. Other fingerprint designs use a combination of the keyed and hashed fingerprinting approaches [152].

One of the attractive advantages of molecular fingerprints is that they are particularly efficient and robust 2D descriptors. In addition, binary representations can be compared extremely quickly due to the fact that they are suited to computer processing. At the same time, there is a significant criticism of all 2D binary representations that concerns the lack of ability to encode the conformer information about a molecule. These representations are obtained only from the connectivity matrix of a compound and thus cannot differentiate between molecular configurations and do not take into account conformational features. More recently some researchers have attempted to involve 3D information into molecular graphs or fingerprints in order to overcome the insufficiency of the 2D descriptors information [153].

3D similarity measures have been reviewed in the literature [3, 153-155]. Examples include field-based approaches, quantum chemical calculations, shape-based approaches and surface-based approaches. Field-based techniques are probably the most popular source of 3D molecular descriptors, and express the 3D molecular field by a collection of sampled data points. At the heart of this approach is the application of a 3D grid, which covers the structures of all the compounds in the database. Quantum chemical calculations characterise molecular features by representations of all the electronic and geometric properties of compounds and their interactions. Shape-based approaches find and calculate the maximal overlap of the volume of two compounds by Gaussian functions [156], spheres [157] or other representations of densities. Surface-based approaches analyse 3D molecular similarity by some representations of the

3D molecular surface. The calculation of similarity based on 3D descriptors is less well established than 2D descriptors because some 3D molecular descriptor based methods require a time-consuming optimisation process and also there is no guarantee they will describe the molecules precisely [158].

1.6.2 Weighting scheme

A weighting scheme is a standardisation method that is applied to normalise different parts of the chosen representation. Thus far, there have been a few studies of the effect of weighting schemes on molecular similarity measures. Willett, et al. [159] reported a systematic comparison of six weighting schemes using six similarity coefficients for the determination of molecular structural similarity. The results suggested that the weighting of fragments by the frequency of occurrence in molecules is more effective than merely noting the presence or absence of a fragment. However, most interest in the chemistry literature has focused on the roles of the representations and of the similarity coefficients in similarity measures rather than the role of weighting schemes [3, 160].

1.6.3 Similarity coefficients

The similarity measures' coefficients or indices are functions that convert pairs of compatible structure representations into numbers. There are many measures to quantify the degree of dissimilarity/similarity between pairs of objects but many of them can be clustered into three broad groups: distance coefficients, correlation coefficients and association coefficients [148, 161, 162]. Association coefficients were originally developed to use with binary data and they are thus very well suited to the measures of fingerprint-based similarities. In most cases the values taken by association coefficients are normalised to lie in the range from 0 to 1. This means that the similarity and dissimilarity coefficients can be converted to each other by subtraction from

1. Todeschini et al. [163] have provided a table that summarises the form and the different properties of 51 similarity coefficients from the literature.

The Tanimoto coefficient is one of the most common similarity measures among the association coefficients. If two molecules B and C have b and c bits set in their fragment bit-strings, with a of these bits being set in both of the fingerprints, then the Tanimoto coefficient is defined to be: $S_{A,B} = \frac{a}{c+b-a}$. The Tanimoto coefficient takes values between zero to unity, with 0 corresponding to no bits in common and 1 to identical fingerprints.

There have been several studies to compare the different association coefficients for fingerprint similarity [164-167]. Holliday et al. [166] carried out a comparison of 22 association coefficients including the Tanimoto index for similarity searching in datasets of chemical compounds represented by 2D fingerprints. This study and an early study from Willett [159] have suggested the use of the Tanimoto coefficient as a superior coefficient for molecular similarity studies. Todeschini et al. [163] reported a comparison of the use of a total of 51 binary similarity coefficients when used for similarity-based searching of some datasets. This study also presented the considerable merits of the well-established Tanimoto coefficient.

1.7 Summary of Project Aims

The overall plan of this project is based on the development of strategies that employ QbD principles to develop chromatographic analytical methods. The research will comprise a series of highly integrated research topics (shown schematically in Figure 1.4) which cover the areas of structure-retention relationships for the HILIC mode, selection of the stationary phases based on predicted retention, and detailed mobile phase optimisation and selection of robust method conditions using QbD principles.

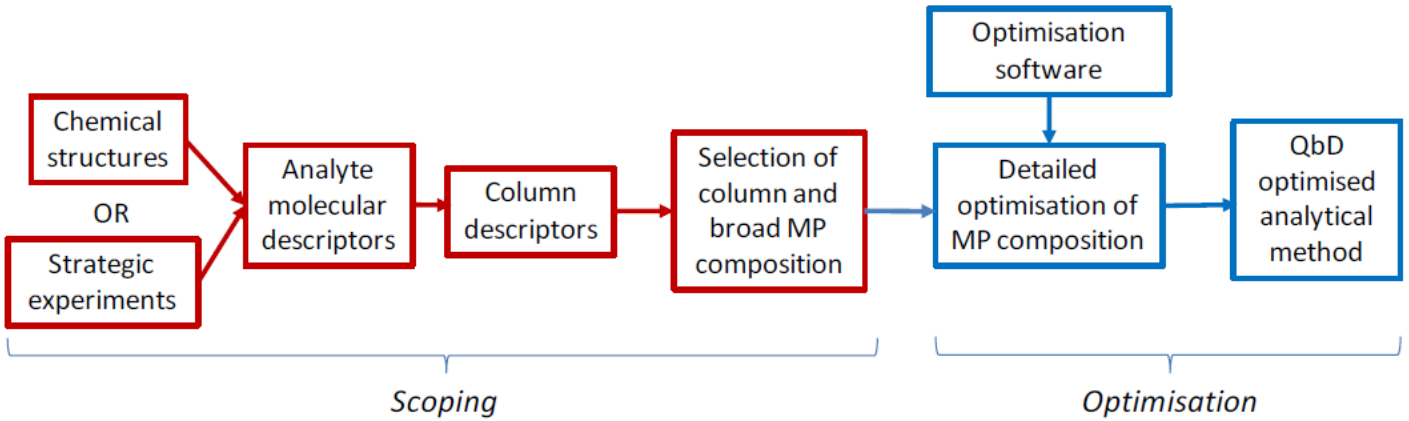

Figure 1.4. Project overview.

Aim 1: To establish a QSRR model based on molecular descriptors computed from chemical structures utilising GA-PLS tools. The prediction of retention allows scoping of HILIC methods for target analytes.

Aim 2: To develop strategies to enhance the accuracy of QSRR methodology in HILIC using a combination of the concepts of structural molecular similarity and chromatographic molecular similarity.

Aim 3: To develop a QbD workflow, which combines DoE principles with QSRR methodology for rapid optimisation of the chromatographic conditions in terms of separation of a mixture of pharmaceutical targets.

1.8 References

[1] B. Buszewski, S. Noga, Hydrophilic interaction liquid chromatography (HILIC)--a powerful separation technique, Anal. Bioanal. Chem., 402 (2012) 231-247.

[2] R. Kaliszan, Quantitative structure-chromatographic retention relationship, Wiley, New York, 1987.

[3] P. Willett, Similarity methods in chemoinformatics, Annu. Rev. Inf. Sci. Technol., 43 (2009) 1-117.

[4] C.A. Doyle, Dorsey, J. G., Reversed-phase HPLC: Preparation and charactrization of reveresed-phase stationary phases, New York, 1998.

[5] A.J. Alpert, Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds, J. Chromatogr. A, 499 (1990) 177-196.

[6] P. Hemström, K. Irgum, Hydrophilic interaction chromatography, J. Sep. Sci., 29 (2006) 1784-1821.

[7] M. Lammerhofer, HILIC and mixed-mode chromatography: the rising stars in separation science, J. Sep. Sci., 33 (2010) 679-680.

[8] Y. Guo, S. Gaiki, Retention behaviour of small polar compounds on polar stationary phases in hydrophilic interaction chromatography, J. Chromatogr. A, 1074 (2005) 71-80.
[9] B. Dejaegher, Y. Vander Heyden, HILIC methods in pharmaceutical analysis, J. Sep. Sci., 33 (2010) 698-715.
[10] A.L. van Nuijs, I. Tarcomnicu, A. Covaci, Application of hydrophilic interaction chromatography for the analysis of polar contaminants in food and environmental samples, J. Chromatogr. A, 1218 (2011) 5964-5974.
[11] T. Tetaz, S. Detzner, A. Friedlein, B. Molitor, J.L. Mary, Hydrophilic interaction chromatography of intact, soluble proteins, J. Chromatogr. A, 1218 (2011) 5892-5896.
[12] H.P. Nguyen, K.A. Schug, The advantages of ESI-MS detection in conjunction with HILIC mode separations: Fundamentals and applications, J. Sep. Sci., 31 (2008) 1465-1480.
[13] D.V. McCalley, Study of the selectivity, retention mechanisms and performance of alternative silica-based stationary phases for separation of ionised solutes in hydrophilic interaction chromatography, J. Chromatogr. A, 1217 (2010) 3408-3417.
[14] P. Jandera, Stationary phases for hydrophilic interaction chromatography, their characterisation and implementation into multidimensional chromatography concepts, J. Sep. Sci., 31 (2008) 1421-1437.
[15] E. Wikberg, T. Sparrman, C. Viklund, T. Jonsson, K. Irgum, A 2H nuclear magnetic resonance study of the state of water in neat silica and zwitterionic stationary phases and its influence on the chromatographic retention characteristics in hydrophilic interaction high-performance liquid chromatography, J. Chromatogr. A, 1218 (2011) 6630-6638.

[16] A.J. Alpert, Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds, J. Chromatogr. A, 499 (1990) 177-196.
[17] Y. Kawachi, T. Ikegami, H. Takubo, Y. Ikegami, M. Miyamoto, N. Tanaka, Chromatographic characterisation of hydrophilic interaction liquid chromatography stationary phases: hydrophilicity, charge effects, structural selectivity, and separation efficiency, J. Chromatogr. A, 1218 (2011) 5903-5919.
[18] N.P. Dinh, T. Jonsson, K. Irgum, Probing the interaction mode in hydrophilic interaction chromatography, J. Chromatogr. A, 1218 (2011) 5880-5891.
[19] A.E. Karatapanis, Y.C. Fiamegos, C.D. Stalikas, A revisit to the retention mechanism of hydrophilic interaction liquid chromatography using model organic compounds, J. Chromatogr. A, 1218 (2011) 2871-2879.
[20] R.I. Chirita, C. West, S. Zubrzycki, A.L. Finaru, C. Elfakir, Investigations on the chromatographic behaviour of zwitterionic stationary phases used in hydrophilic interaction chromatography, J. Chromatogr. A, 1218 (2011) 5939-5963.
[21] A. Kumar, J.C. Heaton, D.V. McCalley, Practical investigation of the factors that affect the selectivity in hydrophilic interaction chromatography, J. Chromatogr. A, 1276 (2013) 33-46.
[22] S. Noga, S. Bocian, B. Buszewski, Hydrophilic interaction liquid chromatography columns classification by effect of solvation and chemometric methods, J. Chromatogr. A, 1278 (2013) 89-97.
[23] A. Periat, B. Debrus, S. Rudaz, D. Guillarme, Screening of the most relevant parameters for method development in ultra-high performance hydrophilic interaction chromatography, J. Chromatogr. A, 1282 (2013) 72-83.

[24] T. Ikegami, K. Tomomatsu, H. Takubo, K. Horie, N. Tanaka, Separation efficiencies in hydrophilic interaction chromatography, J. Chromatogr. A, 1184 (2008) 474-503.
[25] P. Jandera, Stationary and mobile phases in hydrophilic interaction chromatography: a review, Anal. Chim. Acta, 692 (2011) 1-25.
[26] Y. Guo, S. Gaiki, Retention and selectivity of stationary phases for hydrophilic interaction chromatography, J. Chromatogr. A, 1218 (2011) 5920-5938.
[27] G. Greco, T. Letzel, Main interactions and influences of the chromatographic parameters in HILIC separations, J. Chromatogr. Sci., 51 (2013) 684-693.
[28] C.E. Roy, T. Kauss, S. Prevot, P. Barthelemy, K. Gaudin, Analysis of fatty acid samples by hydrophilic interaction liquid chromatography and charged aerosol detector, J. Chromatogr. A, 1383 (2015) 121-126.
[29] A. Petruczynik, M. Waksmundzka-Hajnos, Application of hydrophilic interaction chromatography in phytochemical analysis, Acta Chromatogr., 25 (2013) 1-25.
[30] D.V. McCalley, Is hydrophilic interaction chromatography with silica columns a viable alternative to reversed-phase liquid chromatography for the analysis of ionisable compounds?, J. Chromatogr. A, 1171 (2007) 46-55.
[31] T.L. Chester, Recent developments in high-performance liquid chromatography stationary phases, Anal. Chem., 85 (2013) 579-589.
[32] J.E. Sandoval, J.J. Pesek, Synthesis and characterisation of a hydride-modified porous silica material as an intermediate in the preparation of chemically bonded chromatographic stationary phases, Anal. Chem., 61 (1989) 2067-2075.
[33] J.C. Linden, C.L. Lawhead, Liquid chromatography of saccharides, J. Chromatogr. A, 105 (1975) 125-133.

[34] A.R. Oyler, B.L. Armstrong, J.Y. Cha, M.X. Zhou, Q. Yang, R.I. Robinson, R. Dunphy, D.J. Burinsky, Hydrophilic interaction chromatography on amino-silica phases complements reversed-phase high-performance liquid chromatography and capillary electrophoresis for peptide analysis, J. Chromatogr. A, 724 (1996) 378-383.
[35] F.E. Regnier, R. Noel, Glycerolpropylsilane bonded phases in the steric exclusion chromatography of biological macromolecules, J. Chromatogr. Sci., 14 (1976) 316-320.
[36] X. Wang, W. Li, H.T. Rasmussen, Orthogonal method development using hydrophilic interaction chromatography and reversed-phase high-performance liquid chromatography for the determination of pharmaceuticals and impurities, J. Chromatogr. A, 1083 (2005) 58-62.
[37] P.B. Explore Luna HILIC: Discover HPLC Polar Retention, http://www.phenomenex.com.
[38] Y. Kato, S. Nakatani, T. Kitamura, Y. Yamasaki, T. Hashimoto, Reversed-phase high-performance liquid chromatography of proteins and peptides on a pellicular support based on hydrophilic resin, J. Chromatogr. A, 502 (1990) 416-422.
[39] T. Yoshida, Peptide separation in normal phase liquid chromatography, Anal. Chem., 69 (1997) 3038-3043.
[40] T. Yoshida, Prediction of peptide retention time in normal-phase liquid chromatography, J. Chromatogr. A, 811 (1998) 61-67.
[41] G. Greco, S. Grosse, T. Letzel, Study of the retention behaviour in zwitterionic hydrophilic interaction chromatography of isomeric hydroxy- and aminobenzoic acids, J. Chromatogr. A, 1235 (2012) 60-67.
[42] P. Appelblad, P. Abrahamsson, MS detection of homocysteine, methyl-malonic acid, and succinic acid using HILIC separation on a covalently bonded zwitterionic stationary phase, LCGC North Am., 23 (2005) 24-25.

[43] H. Idborg, L. Zamani, P.O. Edlund, I. Schuppe-Koistinen, S.P. Jacobsson, Metabolic fingerprinting of rat urine by LC/MS Part 1. Analysis by hydrophilic interaction liquid chromatography-electrospray ionization mass spectrometry, J. Chromatogr. B, 828 (2005) 9-13.
[44] K.L. Wade, I.J. Garrard, J.W. Fahey, Improved hydrophilic interaction chromatography method for the identification and quantification of glucosinolates, J. Chromatogr. A, 1154 (2007) 469-472.
[45] R. Oertel, V. Neumeister, W. Kirch, Hydrophilic interaction chromatography combined with tandem-mass spectrometry to determine six aminoglycosides in serum, J. Chromatogr. A, 1058 (2004) 197-201.
[46] Y. Takegawa, K. Deguchi, T. Keira, H. Ito, H. Nakagawa, S. Nishimura, Separation of isomeric 2-aminopyridine derivatized N-glycans and N-glycopeptides of human serum immunoglobulin G by using a zwitterionic type of hydrophilic-interaction chromatography, J. Chromatogr. A, 1113 (2006) 177-181.
[47] N.S. Quiming, N.L. Denola, A.B. Soliev, Y. Saito, K. Jinno, Retention behaviour of ginsenosides on a poly(vinyl alcohol)-bonded stationary phase in hydrophilic interaction chromatography, Anal. Bioanal. Chem., 389 (2007) 1477-1488.
[48] M. Liu, J. Ostovic, E.X. Chen, N. Cauchon, Hydrophilic interaction liquid chromatography with alcohol as a weak eluent, J. Chromatogr. A, 1216 (2009) 2362-2370.
[49] E.S. Grumbach, D.M. Diehl, U.D. Neue, The application of novel 1.7 microm ethylene bridged hybrid particles for hydrophilic interaction chromatography, J. Sep. Sci., 31 (2008) 1511-1518.
[50] M.R. Gama, R.G. da Costa Silva, C.H. Collins, C.B.G. Bottoli, Hydrophilic interaction chromatography, TrAC, Trends Anal. Chem., 37 (2012) 48-60.

[51] J.P. Hutchinson, T. Remenyi, P. Nesterenko, W. Farrell, E. Groeber, R. Szucs, G. Dicinoski, P.R. Haddad, Investigation of polar organic solvents compatible with Corona Charged Aerosol Detection and their use for the determination of sugars by hydrophilic interaction liquid chromatography, Anal. Chim. Acta, 750 (2012) 199-206.
[52] K.J. Fountain, J. Xu, D.M. Diehl, D. Morrison, Influence of stationary phase chemistry and mobile-phase composition on retention, selectivity, and MS response in hydrophilic interaction chromatography, J. Sep. Sci., 33 (2010) 740-751.
[53] C. Dell'Aversano, P. Hess, M.A. Quilliam, Hydrophilic interaction liquid chromatography–mass spectrometry for the analysis of paralytic shellfish poisoning (PSP) toxins, J. Chromatogr. A, 1081 (2005) 190-201.
[54] R.I. Chirita, C. West, A.L. Finaru, C. Elfakir, Approach to hydrophilic interaction chromatography column selection: application to neurotransmitters analysis, J. Chromatogr. A, 1217 (2010) 3091-3104.
[55] M. Isokawa, T. Kanamori, T. Funatsu, M. Tsunoda, Recent advances in hydrophilic interaction chromatography for quantitative analysis of endogenous and pharmaceutical compounds in plasma samples, Bioanalysis, 6 (2014) 2421-2439.
[56] W. Sun, L. Tong, J. Miao, J. Huang, D. Li, Y. Li, H. Xiao, H. Sun, K. Bi, Separation and analysis of phenolic acids from Salvia miltiorrhiza and its related preparations by off-line two-dimensional hydrophilic interaction chromatographyxreversed-phase liquid chromatography coupled with ion trap time-of-flight mass spectrometry, J. Chromatogr. A, 1431 (2016) 79-88.
[57] A.N. Ramdzan, L. Barreiros, M.I. Almeida, S.D. Kolev, M.A. Segundo, Determination of salivary cotinine through solid phase extraction using a bead-injection lab-on-valve approach hyphenated to hydrophilic interaction liquid chromatography, J. Chromatogr. A, 1429 (2016) 284-291.

[58] L. Konieczna, A. Roszkowska, M. Niedzwiecki, T. Baczek, Hydrophilic interaction chromatography combined with dispersive liquid-liquid microextraction as a preconcentration tool for the simultaneous determination of the panel of underivatized neurotransmitters in human urine samples, J. Chromatogr. A, 1431 (2016) 111-121.
[59] E. Cifkova, R. Hajek, M. Lisa, M. Holcapek, Hydrophilic interaction liquid chromatography-mass spectrometry of (lyso)phosphatidic acids, (lyso)phosphatidylserines and other lipid classes, J. Chromatogr. A, 1439 (2016) 65-73.
[60] A. Socia, J.P. Foley, Direct determination of amino acids by hydrophilic interaction liquid chromatography with charged aerosol detection, J. Chromatogr. A, 1446 (2016) 41-49.
[61] A. Periat, S. Fekete, A. Cusumano, J.L. Veuthey, A. Beck, M. Lauber, D. Guillarme, Potential of hydrophilic interaction chromatography for the analytical characterisation of protein biopharmaceuticals, J. Chromatogr. A, 1448 (2016) 81-92.
[62] M.E. Dasenaki, C.S. Michali, N.S. Thomaidis, Analysis of 76 veterinary pharmaceuticals from 13 classes including aminoglycosides in bovine muscle by hydrophilic interaction liquid chromatography-tandem mass spectrometry, J. Chromatogr. A, 1452 (2016) 67-80.
[63] C. Cornell, A. Karanjit, Y. Chen, F. Jacobson, A high-throughput hydrophilic interaction liquid chromatography coupled with a charged aerosol detector method to assess trisulfides in IgG1 monoclonal antibodies using tris(2-carboxyethyl)phosphine reaction products: Tris(2-carboxyethyl)phosphine-oxide and tris(2-carboxyethyl)phosphine-sulfide, J. Chromatogr. A, 1457 (2016) 107-115.
[64] S. Magiera, A. Kolanowska, J. Baranowski, Salting-out assisted extraction method coupled with hydrophilic interaction liquid

chromatography for determination of selected beta-blockers and their metabolites in human urine, J. Chromatogr. B, 1022 (2016) 93-101.
[65] M. Isokawa, T. Shimosawa, T. Funatsu, M. Tsunoda, Determination and characterisation of total thiols in mouse serum samples using hydrophilic interaction liquid chromatography with fluorescence detection and mass spectrometry, J. Chromatogr. B, 1019 (2016) 59-65.
[66] C. Steuer, P. Schutz, L. Bernasconi, A.R. Huber, Simultaneous determination of phosphatidylcholine-derived quaternary ammonium compounds by a LC-MS/MS method in human blood plasma, serum and urine samples, J. Chromatogr. B, 1008 (2016) 206-211.
[67] M. Paczkowska, M. Mizera, A. Tężyk, P. Zalewski, J. Dzitko, J. Cielecka-Piontek, Hydrophilic interaction chromatography (HILIC) for the determination of cetirizine dihydrochloride, Arabian J. Chem., (2016).
[68] G.R. Valicherla, P. Tripathi, S.K. Singh, A.A. Syed, M. Riyazuddin, A. Husain, D. Javia, K.S. Italiya, P.R. Mishra, J.R. Gayen, Pharmacokinetics and bioavailability assessment of Miltefosine in rats using high performance liquid chromatography tandem mass spectrometry, J. Chromatogr. B, 1031 (2016) 123-130.
[69] I.A. Abdallah, D.C. Hammell, A.L. Stinchcomb, H.E. Hassan, A fully validated LC-MS/MS method for simultaneous determination of nicotine and its metabolite cotinine in human serum and its application to a pharmacokinetic study after using nicotine transdermal delivery systems with standard heat application in adult smokers, J. Chromatogr. B, 1020 (2016) 67-77.
[70] J. Maksic, A. Tumpa, A. Stajic, M. Jovanovic, T. Rakic, B. Jancic-Stojanovic, Hydrophilic interaction liquid chromatography in analysis of granisetron HCl and its related substances. Retention mechanisms and method development, J. Pharm. Biomed. Anal., 123 (2016) 93-103.

[71] A. Planinc, B. Dejaegher, Y.V. Heyden, J. Viaene, S. Van Praet, F. Rappez, P. Van Antwerpen, C. Delporte, LC-MS analysis combined with principal component analysis and soft independent modelling by class analogy for a better detection of changes in N-glycosylation profiles of therapeutic glycoproteins, Anal. Bioanal. Chem., (2016).
[72] M. Dousa, R. Klvana, J. Doubsky, J. Srbek, J. Richter, M. Exner, P. Gibala, HILIC-MS determination of Genotoxic impurity of 2-Chloro-N-(2-Chloroethyl)Ethanamine in the vortioxetine manufacturing process, J. Chromatogr. Sci., 54 (2016) 119-124.
[73] R. Joyce, V. Kuziene, X. Zou, X. Wang, F. Pullen, R.L. Loo, Development and validation of an ultra-performance liquid chromatography quadrupole time of flight mass spectrometry method for rapid quantification of free amino acids in human urine, Amino acids, 48 (2016) 219-234.
[74] K. Valkó, L.R. Snyder, J.L. Glajch, Retention in reversed-phase liquid chromatography as a function of mobile-phase composition, J. Chromatogr. A, 656 (1993) 501-520.
[75] P. Nikitas, A. Pappa-Louisi, P. Agrafiotou, Effect of the organic modifier concentration on the retention in reversed-phase liquid chromatography, J. Chromatogr. A, 946 (2002) 33-45.
[76] P.J. Schoenmakers, H.A.H. Billiet, R. Tussen, L. De Galan, Gradient selection in reversed-phase liquid chromatography, J. Chromatogr. A, 149 (1978) 519-537.
[77] U.D. Neue, H.J. Kuss, Improved reversed-phase gradient retention modelling, J. Chromatogr. A, 1217 (2010) 3794-3803.
[78] G. Jin, Z. Guo, F. Zhang, X. Xue, Y. Jin, X. Liang, Study on the retention equation in hydrophilic interaction liquid chromatography, Talanta, 76 (2008) 522-527.

[79] E. Tyteca, A. Periat, S. Rudaz, G. Desmet, D. Guillarme, Retention modelling and method development in hydrophilic interaction chromatography, J. Chromatogr. A, 1337 (2014) 116-127.
[80] R. Kaliszan, QSRR: quantitative structure-(chromatographic) retention relationships, Chem. Rev., 107 (2007) 3212-3246.
[81] K. Jinno, N.S. Quiming, N.L. Denola, Y. Saito, Modelling of retention of adrenoreceptor agonists and antagonists on polar stationary phases in hydrophilic interaction chromatography: a review, Anal. Bioanal. Chem., 393 (2009) 137-153.
[82] K. Gorynski, B. Bojko, A. Nowaczyk, A. Bucinski, J. Pawliszyn, R. Kaliszan, Quantitative structure-retention relationships models for prediction of high performance liquid chromatography retention time of small molecules: endogenous metabolites and banned compounds, Anal. Chim. Acta, 797 (2013) 13-19.
[83] D.J. Creek, A. Jankevics, R. Breitling, D.G. Watson, M.P. Barrett, K.E. Burgess, Toward global metabolomics analysis with hydrophilic interaction liquid chromatography-mass spectrometry: improved metabolite identification by retention time prediction, Anal. Chem., 83 (2011) 8703-8710.
[84] M. Cao, K. Fraser, J. Huege, T. Featonby, S. Rasmussen, C. Jones, Predicting retention time in hydrophilic interaction liquid chromatography mass spectrometry and its use for peak annotation in metabolomics, Metabolomics, 11 (2015) 696-706.
[85] M.H. Abraham, M. Ros□s, C.F. Poole, S.K. Poole, Hydrogen bonding. 42. characterisation of reversed-phase high-performance liquid chromatographic C18 stationary phases, J. Phys. Org. Chem., 10 (1997) 358-368.

[86] S. Studzinska, B. Buszewski, Linear solvation energy relationships in the determination of specificity and selectivity of stationary phases, Chromatographia, 75 (2012) 1235-1246.
[87] G. Schuster, W. Lindner, Comparative characterisation of hydrophilic interaction liquid chromatography columns by linear solvation energy relationships, J. Chromatogr. A, 1273 (2013) 73-94.
[88] L.R. Snyder, J.W. Dolan, P.W. Carr, The hydrophobic-subtraction model of reversed-phase column selectivity, J. Chromatogr. A, 1060 (2004) 77-116.
[89] The United States Pharmacopeial Convention, http://www.usp.org/app/USPNF/columns.html, October 2011.
[90] R. Graesboll, N.J. Nielsen, J.H. Christensen, Using the hydrophobic subtraction model to choose orthogonal columns for online comprehensive two-dimensional liquid chromatography, J. Chromatogr. A, 1326 (2014) 39-46.
[91] J. Wang, Z. Guo, A. Shen, L. Yu, Y. Xiao, X. Xue, X. Zhang, X. Liang, Hydrophilic-subtraction model for the characterisation and comparison of hydrophilic interaction liquid chromatography columns, J. Chromatogr. A, 1398 (2015) 29-46.
[92] M.E. Ibrahim, Y. Liu, C.A. Lucy, A simple graphical representation of selectivity in hydrophilic interaction liquid chromatography, J. Chromatogr. A, 1260 (2012) 126-131.
[93] M. Lammerhofer, M. Richter, J. Wu, R. Nogueira, W. Bicker, W. Lindner, Mixed-mode ion-exchangers and their comparative chromatographic characterisation in reversed-phase and hydrophilic interaction chromatography elution modes, J. Sep. Sci., 31 (2008) 2572-2588.
[94] B. Dejaegher, D. Mangelings, Y. Vander Heyden, Method development for HILIC assays, J. Sep. Sci., 31 (2008) 1438-1448.

[95] K. Monks, I. Molnar, H.J. Rieger, B. Bogati, E. Szabo, Quality by Design: Multidimensional exploration of the design space in high performance liquid chromatography method development for better robustness before validation, J. Chromatogr. A, 1232 (2012) 218-230.
[96] F.G. Vogt, A.S. Kord, Development of quality-by-design analytical methods, J. Pharm. Sci., 100 (2011) 797-812.
[97] E. Rozet, P. Lebrun, P. Hubert, B. Debrus, B. Boulanger, Design spaces for analytical methods, TrAC, Trends Anal. Chem., 42 (2013) 157-167.
[98] G.L. Reid, G. Cheng, D.T. Fortin, J.W. Harwood, J.E. Morgado, J. Wang, G. Xue, Reverse-phase liquid chromatographic method development in an analytical quality by design framework, J. Liq. Chromatogr. Relat. Technol., 36 (2013) 2612-2638.
[99] T. Rakic, B. Jancic Stojanovic, A. Malenovic, D. Ivanovic, M. Medenica, Improved chromatographic response function in HILIC analysis: application to mixture of antidepressants, Talanta, 98 (2012) 54-61.
[100] N.S. Quiming, N.L. Denola, Y. Saito, A.P. Catabay, K. Jinno, Chromatographic behaviour of uric acid and methyl uric acids on a diol column in HILIC, Chromatographia, 67 (2008) 507-515.
[101] M. Fourdinier, S. Bostyn, R. Delepee, H. Fauduet, Interest of a chemometric approach in understanding the retention behaviour of three columns in hydrophilic interaction liquid chromatography: application to the separation of glycerol carbonate, glycerol and urea, Talanta, 81 (2010) 1281-1287.
[102] N. Hatambeygi, G. Abedi, M. Talebi, Method development and validation for optimised separation of salicylic, acetyl salicylic and ascorbic acid in pharmaceutical formulations by hydrophilic interaction chromatography and response surface methodology, J. Chromatogr. A, 1218 (2011) 5995-6003.

[103] K. Novotna, J. Havlis, J. Havel, Optimisation of high performance liquid chromatography separation of neuroprotective peptides. Fractional experimental designs combined with artificial neural networks, J. Chromatogr. A, 1096 (2005) 50-57.
[104] Y. Vander Heyden, C. Perrin, D.L. Massart, Chapter 6 Optimisation strategies for HPLC and CZE, 1 (2000) 163-212.
[105] R. Kaliszan, Quantitative structure-retention relationships applied to reversed-phase high-performance liquid chromatography, J. Chromatogr. A, 656 (1993) 417-435.
[106] R. Put, Y. Vander Heyden, Review on modelling aspects in reversed-phase liquid chromatographic quantitative structure-retention relationships, Anal. Chim. Acta, 602 (2007) 164-172.
[107] R. Todeschini, V. Consonni, Hadbook of molecules descriptors, Wiley, Weinheim, 2000.
[108] in, Talete srl, Dragon 6.0 for Windows (Software For Molecular Descriptor Calculations); http://www.talete.mi.it/ Talete, Milano, Italy [accessed December 2015].
[109] L. Huan, Y. Lei, Toward integrating feature selection algorithms for classification and clustering, IEEE Transactions on Knowledge and Data Engineering, 17 (2005) 491-502.
[110] Y. Saeys, I. Inza, P. Larranaga, A review of feature selection techniques in bioinformatics, Bioinformatics, 23 (2007) 2507-2517.
[111] S. Riahi, M.R. Ganjali, E. Pourbasheer, P. Norouzi, QSRR study of GC retention indices of essential-oil compounds by multiple linear regression with a genetic algorithm, Chromatographia, 67 (2008) 917-922.
[112] J. Li, J. Sun, Z. He, Quantitative structure-retention relationship studies with immobilized artificial membrane chromatography II: partial least squares regression, J. Chromatogr. A, 1140 (2007) 174-179.

[113] C.B. Mazza, N. Sukumar, C.M. Breneman, S.M. Cramer, Prediction of protein retention in ion-exchange systems using molecular descriptors obtained from crystal structure, Anal. Chem., 73 (2001) 5457-5461.
[114] M.H. Fatemi, M.H. Abraham, C.F. Poole, Combination of artificial neural network technique and linear free energy relationship parameters in the prediction of gradient retention times in liquid chromatography, J. Chromatogr. A, 1190 (2008) 241-252.
[115] A.A. D'Archivio, F. Ruggieri, P. Mazzeo, E. Tettamanti, Modelling of retention of pesticides in reversed-phase high-performance liquid chromatography: quantitative structure-retention relationships based on solute quantum-chemical descriptors and experimental (solvatochromic and spin-probe) mobile phase descriptors, Anal. Chim. Acta, 593 (2007) 140-151.
[116] N. Tugcu, M. Song, C.M. Breneman, N. Sukumar, K.P. Bennett, S.M. Cramer, Prediction of the effect of mobile-phase salt type on protein retention and selectivity in anion exchange systems, Anal. Chem., 75 (2003) 3563-3572.
[117] R. Put, C. Perrin, F. Questier, D. Coomans, D.L. Massart, Y. Vander Heyden, Classification and regression tree analysis for molecular descriptor selection and retention prediction in chromatographic quantitative structure–retention relationship studies, J. Chromatogr. A, 988 (2003) 261-276.
[118] T. Hancock, R. Put, D. Coomans, Y. Vander Heyden, Y. Everingham, A performance comparison of modern statistical techniques for molecular descriptor selection and retention prediction in chromatographic QSRR studies, Chemom. Intell. Lab. Syst., 76 (2005) 185-196.
[119] R. Kiralj, M.M.C. Ferreira, Basic validation procedures for regression models in QSAR and QSPR studies: theory and application, J. Braz. Chem. Soc., 20 (2009) 770-787.

[120] K. Varmuza, P. Filzmoser, M. Dehmer, Multivariate linear QSPR/QSAR models: Rigorous evaluation of variable selection for PLS, Comput. Struct. Biotechnol. J., 5 (2013) e201302007.
[121] L. Eriksson, P.L. Andersson, E. Johansson, M. Tysklind, Megavariate analysis of environmental QSAR data. Part II--investigating very complex problem formulations using hierarchical, non-linear and batch-wise extensions of PCA and PLS, Mol. Divers., 10 (2006) 187-205.
[122] L. Eriksson, E. Johansson, F. Lindgren, M. Sjöström, S. Wold, J. Comput.-Aided Mol. Des., 16 (2002) 711-726.
[123] R. Put, M. Daszykowski, T. Baczek, Y. Vander Heyden, Retention prediction of peptides based on uninformative variable elimination by partial least squares, J. Proteome. Res., 5 (2006) 1618-1625.
[124] A. Golbraikh, A. Tropsha, Beware of q2!, J. Mol. Graphics Model., 20 (2002) 269-276.
[125] K. Baumann, N. Stiefl, Validation tools for variable subset regression, J. Comput.-Aided Mol. Des., 18 (2004) 549-562.
[126] D. Baumann, K. Baumann, Reliable estimation of prediction errors for QSAR models under model uncertainty using double cross-validation, J. Cheminform., 6 (2014) 47.
[127] P. Gramatica, Principles of QSAR models validation: internal and external, QSAR Comb. Sci., 26 (2007) 694-701.
[128] R.F. Teófilo, J.P.A. Martins, M.M.C. Ferreira, Sorting variables by using informative vectors as a strategy for feature selection in multivariate regression, J. Chemom., 23 (2009) 32-48.
[129] C. Rucker, G. Rucker, M. Meringer, y-Randomization and its variants in QSPR/QSAR, J. Chem. Inf. Model., 47 (2007) 2345-2357.
[130] C. Wang, M.J. Skibic, R.E. Higgs, I.A. Watson, H. Bui, J. Wang, J.M. Cintron, Evaluating the performances of quantitative structure-retention relationship models with different sets of molecular descriptors and

databases for high-performance liquid chromatography predictions, J. Chromatogr. A, 1216 (2009) 5030-5038.
[131] K. Muteki, J.E. Morgado, G.L. Reid, J. Wang, G. Xue, F.W. Riley, J.W. Harwood, D.T. Fortin, I.J. Miller, Quantitative structure retention relationship models in an analytical quality by design framework: simultaneously accounting for compound properties, mobile-phase conditions, and stationary-phase properties, Ind. Eng. Chem. Res., 52 (2013) 12269-12284.
[132] J.M. A., M.G. M., In Concepts and Applications of Molecular Similarity, John Wiley & Sons, New York, 1990.
[133] R.P. Sheridan, B.P. Feuston, V.N. Maiorov, S.K. Kearsley, Similarity to molecules in the training set is a good discriminator for prediction accuracy in QSAR, J. Chem. Inf. Comput. Sci., 44 (2004) 1912-1928.
[134] H. Yuan, Y. Wang, Y. Cheng, Local and global quantitative structure-activity relationship modelling and prediction for the baseline toxicity, J. Chem. Inf. Model., 47 (2007) 159-169.
[135] C.A. Bergstrom, C.M. Wassvik, U. Norinder, K. Luthman, P. Artursson, Global and local computational models for aqueous solubility prediction of drug-like molecules, J. Chem. Inf. Comput. Sci., 44 (2004) 1477-1488.
[136] H. Zhang, H.Y. Ando, L. Chen, P.H. Lee, On-the-fly selection of a training set for aqueous solubility prediction, Mol. Pharm., 4 (2007) 489-497.
[137] L. He, P.C. Jurs, Assessing the reliability of a QSAR model's predictions, J. Mol. Graph. Model., 23 (2005) 503-523.
[138] A. Koutsoukas, S. Paricharak, W.R. Galloway, D.R. Spring, A.P. Ijzerman, R.C. Glen, D. Marcus, A. Bender, How diverse are diversity assessment methods? A comparative analysis and benchmarking of molecular descriptor space, J. Chem. Inf. Model., 54 (2014) 230-242.

[139] S.L. Dixon, K.M. Merz, One-dimensional molecular representations and similarity calculations: methodology and validation, J. Med. Chem., 44 (2001) 3795-3809.
[140] P. Willett, Similarity searching using 2D structural fingerprints, Methods Mol. Biol., 672 (2011) 133-158.
[141] A.T. Balaban, Chemical graphs: Looking back and glimpsing ahead, J. Chem. Inf. Model., 35 (1995) 339-350.
[142] E. Estrada, E. Uriarte, Curr. Med. Chem., 8 (2001) 1573.
[143] R.P. Sheridan, M.D. Miller, D.J. Underwood, S.K. Kearsley, Chemical similarity using geometric atom pair descriptors, J. Chem. Inf. Model., 36 (1996) 128-136.
[144] S.K. Kearsley, S. Sallamack, E.M. Fluder, J.D. Andose, R.T. Mosley, R.P. Sheridan, Chemical similarity using physiochemical property descriptors, J. Chem. Inf. Model., 36 (1996) 118-127.
[145] C.L. Wilkins, M. Randić, A graph theoretical approach to structure-property and structure-activity correlations, Theor. Chim. Acta, 58 (1980) 45-68.
[146] A.T. Balaban, Highly discriminating distance-based topological index, Chem. Phys. Lett., 89 (1982) 399-404.
[147] P. Willett, Chemoinformatics – similarity and diversity in chemical libraries, Curr. Opin. Biotechnol., 11 (2000) 85-88.
[148] P. Willett, J.M. Barnard, G.M. Downs, Chemical similarity searching, J. Chem. Inf. Model., 38 (1998) 983-996.
[149] J.L. Durant, B.A. Leland, D.R. Henry, J.G. Nourse, Reoptimisation of MDL keys for use in drug discovery, J. Chem. Inf. Model., 42 (2002) 1273-1280.
[150] Daylight Properties Package, Daylight Chemical Information Systems, Inc., http://www.daylight.com.

[151] For details on the hashed ChemAxon fingerprint developed by ChemAxon, see https://docs.chemaxon.com/display/CD/Chemical+Hashed+Fingerprint.
[152] The Unity software packages are available from Tripos Inc at URL: http://www.tripos.com/.
[153] N. Nikolova, J. Jaworska, Approaches to measure chemical similarity– a review, QSAR Comb. Sci., 22 (2003) 1006-1026.
[154] A.G. Maldonado, J.P. Doucet, M. Petitjean, B.T. Fan, Molecular similarity and diversity in chemoinformatics: from theory to applications, Mol. Divers., 10 (2006) 39-79.
[155] W.A. Warr, Representation of chemical structures, Wiley Interdisciplinary Reviews: Computational Molecular Science, 1 (2011) 557-579.
[156] J.A. Grant, B.T. Pickup, A Gaussian description of molecular shape, J. Phys. Chem., 99 (1999) 3503-3510.
[157] M. Petitjean, Geometric molecular similarity from volume-based distance minimization: Application to saxitoxin and tetrodotoxin, J. Comput. Chem., 16 (1995) 80-90.
[158] C. Cai, J. Gong, X. Liu, D. Gao, H. Li, Molecular similarity: methods and performance, Chin. J. Chem., 31 (2013) 1123-1132.
[159] P. Willett, V. Winterman, A comparison of some measures for the determination of inter-molecular structural similarity measures of inter-molecular structural similarity, Quant. Struct.-Act. Relat., 5 (1986) 18-25.
[160] P. Willett, Similarity-based data mining in files of two-dimensional chemical structures using fingerprint measures of molecular resemblance, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1 (2011) 241-251.
[161] G.U. Yule, On the association of attributes in statistics., Philos. Trans. R. Soc. A, 75 257-319.

[162] K. Pearson, D. Heron, On theories of association, Biometrika, 9 (1913) 159-315.
[163] R. Todeschini, V. Consonni, H. Xiang, J. Holliday, M. Buscema, P. Willett, Similarity coefficients for binary chemoinformatics data: overview and extended comparison using simulated and real data sets, J. Chem. Inf. Model., 52 (2012) 2884-2901.
[164] A. Bender, J.L. Jenkins, J. Scheiber, S.C. Sukuru, M. Glick, J.W. Davies, How similar are similarity searching methods? A principal component analysis of molecular descriptor space, J. Chem. Inf. Model., 49 (2009) 108-119.
[165] J. Duan, S.L. Dixon, J.F. Lowrie, W. Sherman, Analysis and comparison of 2D fingerprints: insights into database screening performance using eight fingerprint methods, J. Mol. Graph. Model., 29 (2010) 157-170.
[166] Holliday, Grouping of coefficients for the calculation of inter-molecular similarity and aissimilarity using 2D fragment bit-strings, Comb. Chem. High Throughput Screening, 5 (2002).
[167] M. Sastry, J.F. Lowrie, S.L. Dixon, W. Sherman, Large-scale systematic analysis of 2D fingerprint methods and parameters to improve virtual screening enrichments, J. Chem. Inf. Model., 50 (2010) 771-784.

2 Experimental section and data collection

2.1 Data collection

2.1.1 Sample preparation

Standards of adrenaline, noradrenaline, isoproterenol, salbutamol, dopamine, tyramine, synephrine, 3-methoxytyramine, norfenefrine, normetanephrine, N-methylephedrine, octopamine, salicylic acid, 4-hydroxybenzoic acid, 2,3-dihydroxybenzoic acid, 2,4-dihydroxybenzoic acid, 3,5-dihydroxybenzoic acid, benzoic acid, 3-amino-4-hydroxybenzoic acid, 3-aminobenzoic acid, vanillic acid, syringic acid, 2-methoxybenzoic acid, p-toluic acid, 3-hydroxybenzoic acid, 2,5-dihydroxybenzoic acid, 3,4-dihydroxybenzoic acid, 4-aminobenzoic acid, 4-aminosalicylic acid, 2'-deoxyadenosine, 2'-deoxycytidine, 2'-deoxyguanosine, adenosine, cytidine, guanosine, inosine, thymidine, uridine, 2'-deoxyuridine, acyclovir, guanine, xanthine, caffeine, theophylline, theobromine, hypoxanthine, 1,3-dimethyluric acid, 5-sulfosalicylic acid, mandelic acid, nicotinic acid, p-toluenesulfonic acid, 4-hydroxybenzenesulfonic acid, tropic acid, 4-aminophenylacetic acid, tryptophan, 2-phenylethylamine, phenylalanine, benzyltrimethylammonium (BTMA) chloride, phenyltrimethylammonium (PTMA) chloride, labetalol, nadolol, propranolol, adenine, uracil, thymine, cytosine, pindolol, alprenolol, satolol, atenolol, 4-nitrophenyl-β-D-glycopyranoside, uric acid, vidarabine, gallic acid, phthalic acid, isophthalic acid, terephthalic acid, caffeic acid, 3,5-dinitrosalicylic acid, acetylsalicylic acid, phenylacetic acid, 1,2-benzenedisulfonic acid, malic acid, 2-sulfobenzoic acid, ceftiofur and tyrosine were purchased from Sigma–Aldrich (St. Louis, MO, USA) and the standards of fenotrole, ritudrine, metaproterenol, isoxuprine, terbutaline, phenylephrine, methoxamine, 5-methylsalicylic acid, 2',3'-dideoxyadenosine, 3'-deoxyguanosine, 5-methyluridine, 3'-deoxythymidine, 2'-deoxyinosine, 7-methylguanosine,

pentoxyphylline, diphylline, 7-hydroxyethyl-theophylline, 1-methyluric acid, 1-methyl-guanine, 9-methyl-guanine, 3,7-dimethyluric acid, 7-methyl-xanthine, 1,7-dimethyluric acid, proxyphylline and 1,3,7-trimethyl uric acid were purchased from Santa Cruz Biotechnology Inc. (CA, USA). Acetonitrile and methanol of HPLC grade were supplied by VWR International (Melbourne, VIC, Australia) and Sigma–Aldrich (St. Louis, MO, USA), respectively. Formic acid (FA), ammonium formate (NH_4FA), and ammonium hydroxide (NH_4OH), all of analytical grade, were obtained from Sigma–Aldrich (St. Louis, MO, USA). 18.2 MΩ Milli-Q water produced using a Millipore Gradient water purification (Millipore, Bedford, MA, USA) system, was used to prepare mobile phase and sample solutions.

2.1.2 Standard solutions

Standard stock solutions (1000 μg mL^{-1}) of each analyte were obtained by dissolving an appropriate amount of each standard with the appropriate solvent. For β-adrenergic agonists and β-blockers methanol was used, with the exception of adrenaline and 3-methoxytyramine, which were prepared in acidified methanol (0.5% 1M formic acid in methanol); for benzoic acids and nucleosides the standard solutions were prepared in acetonitrile-water (90:10) solution. 0.01M NaOH solution was used to dissolve the uric acids, and the standard solutions of the xanthines were prepared in water, with the exception of 1-methyl-guanine and guanine which were prepared in aqueous formic acid (1% v/v). The aqueous solutions of guanine, xanthine, 7-methyl-xanthine, theobromine, hypoxanthine, 1-methyluric acid, uric acid and vidarabine were centrifuged and the supernatant used as the stock solution. The standard solutions for the rest of the compounds were prepared in acetonitrile-water (50:50) solution. The individual stock solutions were stored at –20 °C and were stable for at least 3 months. A working standard

(100 μg mL^{-1}) of each compound was prepared by diluting the stock solution in acetonitrile: Milli-Q water (90:10).

2.1.3 Instrumentation

All experiments were performed using a Thermo Fisher Scientific Ultimate 3000 instrument (Lane Cove, Australia) equipped with a DGP-3600RS pump, a DAD-3000RS diode array detector, a WPS-3000TRS autosampler with temperature control, and a TCC-3000RS column compartment. Chromeleon software (ver. 7.1.2) was used for system control and data processing. The HPLC columns employed in this study were all obtained from Thermo Fisher Scientific and consisted of a bare silica (Accucore, 4.6 mm id × 150 mm, 2.6 μm); an amino (Syncronis, 4.6 mm id × 150 mm, 3.0 μm); an amide (Acclaim HILIC-10, 3.0 mm id × 150 mm, 3.0 μm), a diol (Acclaim Mixed-Mode HILIC-1, 3.0 mm id × 150 mm, 3.0 μm, dual reversed phase/HILIC mechanism), and a zwitterionic column (Syncronis HILIC, 4.6 mm id × 150 mm, 3.0 μm). In this thesis, these five different columns are referred to as bare silica, amine, amide, diol and zwitterionic, respectively. The isocratic eluents used contained acetonitrile−formate buffer solution. Formate buffer was prepared with an adapted volume of ammonium formate and the pH adjusted to the desired value with formic acid or ammonium hydroxide. The pH measurements were performed at 25°C with a TPS pH meter (QLD, Australia), before the addition of the organic solvent.

All data were collected at a column temperature of 25°C at an eluent flow-rate of 1.0 mL/min for the zwitterionic and amine columns, 0.4 mL/min for the diol and amide columns, and 1.5 mL/min for the bare silica column. Columns were equilibrated with 15–20 column volumes of eluent to guarantee stable equilibrium situations. The void time of each column was measured by the injection of acetone [1]. The typical injection volume

was 5 µL. UV detection was carried out at 254 nm and 280 nm to obtain maximal absorbance for all analytes.

2.1.4 Design of experiments

The stationary phase type, pH, buffer concentration, and acetonitrile content have been proven to be the critically influential parameters for tuning selectivity in the HILIC mode [2-4]. Consequently, a central composite design for three selected chromatographic factors (acetonitrile concentration, pH and salt concentration) on the amide column and a full factorial design on the bare silica, amine, diol, and zwitterionic stationary phases were carried out. Table 2.1 shows the levels studied for selected critical chromatographic parameters. Design matrices with 17 independent trials for the central composite design, and 11 for the full factorial design, were constructed as seen in Tables 2.1-2.3 and Figure 2.1.

The retention times on the amide, amine, zwitterionic, bare silica, and diol stationary phases are provided in Tables 4-8, respectively.

2.2 Model generation

2.2.1 Software

MarvinSketch version 16.2.15 from ChemAxon [5] (Budapest, Hungary) was used for drawing the molecular structures. Initial conformational searches to find the 50 lowest energy structures were performed using a Merck Molecular Force Field (MMff94) [6-9] implemented in Balloon [10]. Geometry optimisation using the semi-empirical Parametric Method number 7 (PM7) [11] was performed in Molecular Orbital PACkage (MOPAC) [12], followed by further geometry optimisation of the lowest energy structure using density functional theory [13, 14] implemented in Gaussian 09 [15]. Finally, Dragon 6.0 software [16] (Talete, Milano, Italy) was employed for calculation of molecular descriptors. A genetic algorithm in Matlab [17]

Table 2.1. Experimental design to optimise the HILIC method development assay

factor	low (-1)	central (0)	high (1)
Acetonitrile content (%) in MP	70	80	90
pH in water phase	3	5	7
Salt concentration (mM) in MP	10	15	20

MP is the mobile phase

Table 2.2. Plan of experiments defined by the central composite design.

nr.	Acentonitrile content in mobile phase	pH	Salt concentration in mobile phase
1	70	3	10
2	90	3	10
3	70	7	10
4	90	7	10
5	70	3	20
6	90	3	20
7	70	7	20
8	90	7	20
9	70	5	15
10	90	5	15
11	80	3	15
12	80	7	15
13	80	5	10
14	80	5	20
15	80	5	15
16	80	5	15
17	80	5	15

Table 2.3. Plan of experiments defined by the full factorial design.

nr.	Acentonitrile content in mobile phase	pH	Salt concentration in mobile phase
1	70	3	10
2	90	3	10
3	70	7	10
4	90	7	10
5	70	3	20
6	90	3	20
7	70	7	20
8	90	7	20
9	80	5	15
10	80	5	15
11	80	5	15

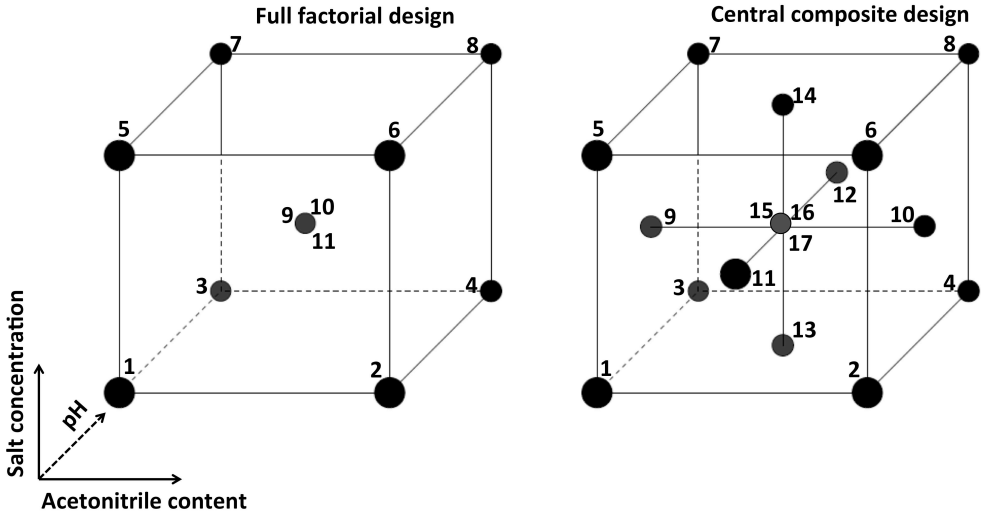

Figure 2.1. Diagram of the design space.

Table 2.4. Experimentally obtained retention times for each operating condition of the central composite design using amide column.

Analyte	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Acetone	2.00	2.04	2.01	2.04	2.01	2.03	2.01	2.03	2.01	2.04	2.02	2.02	2.02	2.01	2.06	2.01	2.02
Toluene	1.83	1.89	1.84	1.88	1.85	1.87	1.83	1.87	1.84	1.87	1.84	1.84	1.85	1.82	1.83	1.83	1.84
2'-Deoxyadenosine	2.60	4.73	2.60	4.71	2.60	4.93	2.63	4.99	2.61	4.75	3.16	3.15	3.14	3.19	3.16	3.15	3.14
2'-Deoxycytidine	2.87	10.49	3.06	10.47	2.92	10.75	3.11	11.58	3.07	10.46	4.24	4.35	4.35	4.42	4.36	4.34	4.33
2',3'-Dideoxyadenosine	2.54	3.98	2.55	3.96	2.55	4.03	2.57	4.07	2.56	3.95	2.96	2.95	2.95	2.98	2.95	2.94	2.94
2'-Deoxyguanosine	2.89	8.93	2.88	8.75	2.89	9.88	2.92	10.10	2.89	9.08	3.95	3.95	3.93	4.01	3.95	3.93	3.93
3'-Deoxyguanosine	2.87	8.70	2.87	8.51	2.87	9.74	2.90	9.79	2.87	8.81	3.87	3.86	3.86	3.93	3.88	3.86	3.86
5'-Methyluridine	2.46	4.16	2.45	4.13	2.46	4.58	2.48	4.52	2.46	4.27	2.89	2.88	2.88	2.93	2.90	2.89	2.88
8-Hydroxy-2'-deoxyguanosine	2.92	9.50	2.93	9.30	2.92	10.79	2.97	10.80	2.93	9.63	4.02	4.03	4.02	4.10	4.05	4.01	4.01
Adenosine	2.68	5.46	2.67	5.42	2.68	5.99	2.71	5.93	2.68	5.55	3.35	3.33	3.33	3.40	3.35	3.34	3.33
Cytidine	2.98	12.00	3.12	11.90	3.02	13.45	3.17	13.71	3.12	12.17	4.51	4.56	4.57	4.66	4.59	4.56	4.56
Guanosine	2.99	10.91	2.99	10.70	2.99	12.84	3.03	12.67	2.99	11.22	4.26	4.26	4.24	4.34	4.28	4.25	4.25
Inosine	2.72	6.91	2.73	6.91	2.73	8.12	2.76	8.09	2.73	7.29	3.57	3.59	3.56	3.65	3.59	3.58	3.58
Thymidine	2.38	3.53	2.37	3.51	2.38	3.77	2.39	3.73	2.38	3.58	2.71	2.70	2.69	2.73	2.70	2.70	2.69
Uridine	2.48	4.30	2.47	4.27	2.48	4.85	2.50	4.76	2.48	4.46	2.95	2.95	2.93	3.00	2.96	2.95	2.94
Adrenaline	2.00	7.75	2.56	9.38	2.28	8.33	2.65	10.39	2.47	9.12	3.09	3.69	3.45	3.58	3.48	3.45	3.50
Noradrenaline	2.05	9.81	2.63	12.26	2.35	10.86	2.73	13.45	2.54	11.32	3.32	3.96	3.70	3.85	3.75	3.71	3.77
3'-Methoxytyramine	1.92	5.98	2.39	6.55	2.17	6.16	2.47	7.19	2.33	6.64	2.78	3.21	2.63	3.15	3.07	3.04	3.07
Isoproterenol	1.93	6.40	2.44	-	2.21	6.66	2.54	8.45	2.36	7.20	2.86	3.33	3.13	3.24	3.15	3.15	3.19
Fenotrole	1.91	6.75	2.33	6.83	2.15	6.83	2.41	7.42	2.30	6.85	2.79	3.13	3.00	3.07	3.01	2.98	3.01
Terbutaline	1.91	5.88	2.32	6.25	2.15	5.98	2.40	6.70	2.28	6.22	2.74	3.10	2.96	3.04	2.97	2.95	2.98

Salbutamol	1.93	6.49	2.39	6.99	2.20	6.68	2.48	7.62	2.35	6.99	2.86	3.29	3.12	3.22	3.13	3.11	3.14
Ritudrine	1.84	5.43	2.25	5.66	2.08	5.37	2.33	6.01	2.22	5.63	2.57	2.90	2.78	2.84	2.79	2.76	2.79
Metaproterenol	1.94	6.41	2.37	6.86	2.18	6.62	2.46	7.46	2.33	6.88	2.84	3.24	3.08	3.18	3.09	3.07	3.10
Synephrine	1.94	6.32	2.43	7.10	2.19	6.53	2.52	7.74	2.36	7.15	2.85	3.36	3.16	3.27	3.17	3.15	3.19
Dopamine	2.02	7.64	2.57	8.38	2.28	8.16	2.61	9.54	2.48	8.60	3.09	3.52	3.43	3.49	3.41	3.39	3.43
N-methylephrine	1.81	3.98	2.25	4.34	2.04	3.89	2.31	4.40	2.19	4.26	2.43	2.79	2.66	2.71	2.65	2.64	2.67
Norphenylephrine	1.99	7.58	2.48	8.12	2.25	7.99	2.57	9.06	2.41	8.23	3.02	3.51	3.31	3.43	3.34	3.31	3.35
Phenylephrine	1.94	6.06	2.41	6.75	2.18	6.24	2.50	7.28	2.35	6.73	2.82	3.30	3.11	3.21	3.12	3.10	3.14
Tyramine	1.92	6.26	2.39	6.90	2.17	6.46	2.47	7.60	2.33	6.98	2.79	3.25	3.07	3.17	3.09	3.07	3.10
Normetanephrine	1.98	7.55	2.49	8.24	2.26	7.97	2.58	9.18	2.43	8.31	3.04	3.55	3.35	3.47	3.37	3.35	3.39
Octopamine	1.99	7.93	2.50	8.63	2.26	8.40	2.59	9.75	2.44	8.78	3.06	3.58	3.38	3.50	3.40	3.38	3.42
Methoxamine	1.86	5.04	2.27	5.39	2.09	5.01	2.34	5.67	2.23	5.33	2.58	2.92	2.79	2.86	2.80	2.79	2.81
Isoxuprine	1.76	3.98	2.10	3.46	1.97	3.84	2.15	3.51	2.08	3.48	2.32	2.45	2.42	2.45	2.42	2.42	2.42
Ceftiofur	4.86	9.11	3.80	9.14	3.62	8.08	3.36	9.15	3.67	8.37	4.14	4.10	4.53	4.04	4.32	4.32	4.25
Tyrosine	3.22	24.42	3.32	24.33	3.24	25.85	3.33	29.55	3.28	24.50	5.26	5.46	5.44	5.41	5.41	5.40	5.44
Pentoxyfylline	2.12	2.37	2.12	2.36	2.12	2.36	2.12	2.36	2.13	2.35	2.21	2.21	2.21	2.20	2.20	2.22	2.20
Guanine	3.04	9.73	3.07	9.35	3.06	10.58	3.10	10.94	3.07	9.63	4.15	4.21	4.17	4.21	4.18	4.18	4.18
Xanthine	2.65	5.39	2.78	5.60	2.67	5.85	2.77	6.41	2.71	5.73	3.23	3.40	3.33	3.35	3.34	3.34	3.34
Caffeine	2.24	2.50	2.24	2.46	2.25	2.52	2.25	2.53	2.25	2.49	2.35	2.35	2.35	2.35	2.34	2.36	2.35
Theophylline	2.34	3.02	2.36	2.98	2.35	3.05	2.37	3.07	2.36	3.00	2.54	2.56	2.55	2.56	2.55	2.57	2.55
Theobromine	2.39	3.11	2.40	3.05	2.40	3.15	2.41	3.16	2.40	3.07	2.63	2.63	2.62	2.63	2.62	2.64	2.62
Diphylline	2.39	3.50	2.40	3.41	2.41	3.64	2.41	3.66	2.41	3.47	2.72	2.73	2.71	2.74	2.72	2.74	2.72
7-Hydroxyethyltheophylline	2.31	2.92	2.31	2.87	2.32	2.91	2.32	2.93	2.32	2.88	2.51	2.52	2.51	2.52	2.51	2.53	2.51

1-Methyluric acid	2.86	7.36	4.70	17.37	2.82	7.37	4.21	19.68	4.11	15.80	3.57	6.16	6.40	5.73	6.17	6.02	6.07
1-Methylguanine	2.77	6.18	2.80	5.90	2.80	6.43	2.82	6.63	2.81	6.07	3.48	3.51	3.49	3.52	3.49	3.51	3.50
9-Methylguanine	2.83	6.57	2.85	6.25	2.86	6.89	2.88	7.09	2.86	6.44	3.59	3.61	3.58	3.61	3.59	3.61	3.60
Uric acid	3.27	13.83	5.56	34.38	3.17	13.56	4.86	41.13	4.98	31.95	4.51	8.20	8.87	7.73	8.60	8.46	-
3,7-Dimethyluric acid	2.66	5.21	3.86	9.64	2.65	5.33	3.54	11.67	3.43	9.62	3.16	4.70	4.69	4.38	4.64	4.56	4.57
7-Methylxanthine	2.50	4.03	2.53	3.92	2.53	4.17	2.55	4.28	2.53	4.01	2.90	2.92	2.90	2.92	2.91	2.93	2.91
Hypoxanthine	2.65	5.52	2.70	5.36	2.68	5.89	2.71	6.20	2.69	5.59	3.28	3.34	3.29	3.33	3.31	3.33	3.32
Proxyphylline	2.26	2.75	2.26	2.71	2.27	2.69	2.27	2.74	2.27	2.64	2.42	2.43	2.42	2.43	2.43	2.45	2.43
1,7-Dimethyluric acid	2.70	5.27	4.36	11.69	2.66	5.11	3.92	12.98	3.80	10.47	3.17	5.36	5.41	4.88	5.32	5.20	5.21
1,3-Dimethyluric acid	2.57	4.72	3.47	7.53	2.57	4.85	3.23	8.70	3.07	7.20	3.01	4.13	4.06	3.78	4.00	3.91	3.92
1,3,7-Trimethyluric acid	2.30	2.81	2.32	2.75	2.32	2.71	2.32	2.76	2.40	3.08	2.61	2.62	2.61	2.62	2.47	2.51	2.48
4-Nitrophenyl-B-D-glycopyranoside	2.25	3.48	2.25	3.38	2.25	3.67	2.25	3.76	2.25	3.49	2.52	2.52	2.52	2.53	2.52	2.55	2.52
Acyclovir	2.78	7.60	2.79	7.18	2.80	8.18	2.82	8.77	2.80	7.58	3.66	3.70	3.65	3.71	3.70	3.73	3.69
2'-Deoxyuridine	2.39	3.73	2.40	3.62	2.41	3.91	2.41	4.06	2.40	3.76	2.77	2.77	2.75	2.79	2.78	2.82	2.77
3'-Deoxythymidine	2.28	2.95	2.29	2.90	2.29	2.97	2.30	3.01	2.29	2.92	2.48	2.48	2.48	2.49	2.49	2.52	2.48
2'-Deoxyinosine	2.63	5.95	2.65	5.75	2.66	6.45	2.67	6.96	2.65	6.09	3.35	3.39	3.34	3.40	3.39	3.43	3.38
Satolol	1.85	4.76	2.25	5.09	2.06	4.70	2.31	5.32	2.22	5.03	2.55	2.86	2.74	2.80	2.75	2.78	2.76
7-Methylguanosine	2.20	15.24	3.16	19.93	2.55	16.29	3.23	24.69	2.92	19.77	4.04	5.51	5.01	5.15	5.05	5.01	5.07
Atenolol	1.97	7.45	2.47	8.11	2.25	7.45	2.56	9.05	2.41	8.12	3.02	3.53	3.32	3.44	3.36	3.39	3.37
Vadarabine	2.70	6.34	2.72	6.13	2.73	6.76	2.74	7.28	2.73	6.37	3.49	3.50	3.47	3.52	3.52	3.57	3.50
Tryptophan	3.27	19.08	3.36	18.99	3.28	18.95	3.37	22.16	3.33	18.71	4.95	5.09	5.11	5.05	5.14	5.18	5.10
BTMA	1.80	3.64	2.21	4.02	2.02	3.54	2.29	4.10	2.17	4.02	2.38	2.70	2.58	2.65	2.58	2.63	2.61

PTMA	1.83	3.92	2.29	4.39	2.07	3.84	2.37	4.53	2.24	4.42	2.47	2.86	2.71	2.80	2.71	2.76	2.75
Labetalol	1.84	4.77	2.28	5.00	2.06	4.58	2.34	5.00	2.22	4.81	2.52	2.83	2.73	2.76	2.73	2.76	2.73
Nadolol	1.97	7.17	2.47	7.85	2.25	7.20	2.56	8.69	2.41	7.71	2.99	3.45	3.27	3.38	3.31	3.35	3.32
Propranolol	1.79	3.93	2.21	4.20	1.99	3.77	2.25	4.16	2.14	4.05	2.37	2.63	2.53	2.57	2.54	2.58	2.55
Adenine	2.73	5.32	2.79	5.31	2.77	5.45	2.81	5.82	2.79	5.32	3.38	3.38	3.38	3.40	3.40	3.48	3.40
Uracil	2.38	3.27	2.39	3.27	2.39	3.38	2.40	3.51	2.39	3.32	2.65	2.65	2.65	2.67	2.66	2.73	2.66
Thymine	2.37	3.18	2.37	3.18	2.37	3.25	2.38	3.35	2.37	3.20	2.60	2.60	2.60	2.61	2.61	2.68	2.61
Cytosine	2.89	10.91	3.40	11.30	2.99	10.50	3.43	12.39	3.38	10.92	4.39	4.77	4.81	4.78	4.87	4.96	4.83
Pindolol	1.81	4.21	2.21	4.52	2.01	4.09	2.27	4.56	2.18	4.41	2.43	2.71	2.61	2.66	2.61	2.67	2.62
Alprenolol	1.74	3.54	2.09	3.75	1.92	3.41	2.17	3.71	2.05	3.64	2.33	2.48	2.42	2.43	2.41	2.46	2.41
Salicylic acid	4.02	3.70	3.13	3.60	3.09	3.51	2.80	3.60	3.06	3.50	3.05	2.88	3.15	2.88	3.09	3.18	3.04
5-Methylsalicylic acid	3.71	3.58	3.09	3.57	2.93	3.36	2.78	3.54	3.02	3.45	2.89	2.83	3.10	2.84	3.04	3.13	2.99
4-Hydroxybenzoic acid	2.19	2.75	4.06	6.27	2.17	2.70	3.59	6.19	3.31	5.41	2.25	3.96	3.89	3.54	3.91	3.85	3.78
3-Hydroxybenzoic acid	2.34	3.09	3.99	7.50	2.25	2.97	3.52	7.36	3.58	6.67	2.34	4.12	4.37	3.85	4.33	4.35	4.21
2,3-Dihydroxybenzoic acid	4.62	4.71	3.46	4.54	3.49	4.54	3.11	4.50	3.36	4.39	3.57	3.27	3.62	3.29	3.57	3.69	3.51
2,4-Dihydroxybenzoic acid	3.51	4.60	3.44	5.18	2.89	4.29	3.09	5.14	3.35	4.99	3.04	3.38	3.73	3.38	3.67	3.77	3.60
2,5-Dihydroxybenzoic acid	4.17	4.77	3.20	4.59	3.22	4.62	2.89	4.64	3.14	4.46	3.38	3.15	3.45	3.17	3.39	3.50	3.34
3,4-Dihydroxybenzoic acid	2.33	3.23	4.42	-	2.28	3.25	3.88	-	3.56	7.59	2.42	3.15	4.47	4.04	4.52	4.23	4.39
3,5-Dihydroxybenzoic acid	2.43	3.78	3.88	9.69	2.32	3.69	3.45	9.80	3.55	8.54	2.50	4.31	4.61	4.13	4.55	4.34	4.44
Benzoic acid	2.29	2.68	4.05	6.00	2.19	2.54	3.57	5.64	3.58	5.15	2.21	3.84	4.14	3.61	4.11	3.94	-
3-Amino-4-hydroxybenzoic acid	2.31	3.04	4.39	7.69	2.27	3.03	3.88	7.63	3.50	6.41	2.40	4.48	4.34	3.92	4.38	4.06	4.24
Gallic acid	2.46	4.20	5.13	-	2.41	4.39	4.60	-	3.86	-	2.65	-	5.31	4.72	6.98	-	5.54

4-Aminobenzoic acid	2.14	2.47	3.71	4.34	2.12	2.48	3.36	4.27	2.92	3.78	2.20	3.39	3.23	3.01	3.25	3.07	3.15
4-Aminosalicylic acid	2.80	3.59	3.66	5.40	2.50	3.32	3.27	5.24	3.52	5.01	2.58	3.58	3.96	3.55	3.89	3.91	3.81
Phthalic acid	4.96	3.83	11.60	10.35	3.49	3.77	7.72	11.14	5.90	7.70	3.35	6.01	6.48	5.01	7.29	6.41	6.28
Isophthalic acid	4.07	9.45	18.17	82.07	3.11	7.69	10.19	72.92	12.55	58.52	3.47	15.54	22.79	13.36	20.91	19.06	18.64
3-Aminobenzoic acid	2.33	2.82	4.34	7.51	2.24	2.72	3.82	7.25	3.76	6.25	2.31	4.40	4.60	4.09	4.61	4.32	4.44
Vanillic acid	2.22	2.76	4.18	6.71	2.19	2.69	3.71	6.55	3.48	5.59	2.27	4.13	4.17	3.72	4.18	3.91	4.07
Syringic acid	2.30	2.89	4.26	7.57	2.23	2.80	3.78	7.33	3.66	6.29	2.31	4.35	4.50	4.01	4.50	4.20	4.34
2-Methoxybenzoic acid	2.32	2.69	4.67	7.63	2.22	2.56	4.06	7.27	3.97	6.17	2.23	4.48	4.76	4.14	4.80	4.47	4.55
Terephthalic acid	4.53	-	21.49	112.35	3.33	8.18	11.96	97.28	14.02	74.27	3.68	19.47	27.69	16.36	26.60	23.62	23.67
Caffeic acid	2.32	3.17	4.70	9.62	2.25	3.15	4.11	-	3.72	7.60	2.38	4.66	4.65	4.15	4.75	4.36	4.52
P-Toluic acid	2.12	2.39	3.85	5.17	2.07	2.35	3.42	5.02	3.31	4.46	2.11	3.55	3.67	3.29	3.68	3.48	3.55
3,5-Dinitrosalicylic acid	2.91	1.88	2.21	1.80	2.36	1.90	2.06	1.81	2.19	1.82	2.10	1.94	2.02	1.93	1.97	2.00	1.98
Nicotinic acid	4.47	8.72	4.96	14.15	3.59	7.32	4.33	13.52	4.78	12.31	4.25	5.68	6.72	5.67	6.65	6.31	6.36
4-Hydroxybenzenesulfonic acid	4.92	5.85	3.25	5.16	3.60	6.02	2.96	5.41	3.18	5.05	3.98	3.30	3.63	3.33	3.58	3.53	3.53
4-Aminophenylacetic acid	2.32	2.83	5.00	8.54	2.25	2.58	4.36	8.51	4.00	6.79	2.33	4.90	5.03	4.39	5.29	4.72	4.89
Acetylsalicylic acid	-	2.99	-	-	-	-	-	-	3.39	5.23	2.31	3.57	3.93	3.49	-	-	-
Phenylacetic acid	2.16	-	4.31	6.14	2.09	-	3.79	6.08	3.62	5.12	2.13	-	4.20	3.65	4.30	3.98	4.12
1,2-benzenedisulfonic acid	-	-		-	28.84	-	11.77	-	16.40	-	24.19	12.06	21.25	10.71	-	17.77	16.38
P-Toluenesulfonic acid	4.49	3.97	3.01	3.58	3.32	3.97	2.72	3.62	2.95	3.49	3.32	2.82	3.08	2.85	3.04	3.01	2.99
2-Sulfobenzoic acid	20.04	17.68	-	47.06	8.06	15.80	11.96	40.51	17.02	35.25	7.87	13.98	23.96	12.12	-	19.51	18.03
Tropic acid	2.37	3.25	3.99	8.04	2.26	3.09	3.54	8.15	3.63	6.98	2.37	4.19	4.55	4.07	4.41	4.28	4.36
2-Phenylethylamine	1.86	4.88	2.31	5.34	2.09	4.86	2.38	5.42	2.24	5.20	2.57	2.94	2.80	2.88	2.82	2.72	2.83
Phenylalanine	3.11	16.39	3.18	16.36	3.11	16.56	3.21	17.64	3.15	15.60	4.60	4.70	4.76	4.72	4.82	4.52	4.76

Mandelic acid	3.89	5.64	3.96	6.89	3.15	5.09	3.52	6.76	3.81	6.29	3.39	-	4.54	4.06	4.52	4.38	4.36
5-Sulfosalicylic acid	27.58	19.06	7.51	12.05	9.51	17.54	4.97	14.56	6.41	13.09	9.85	5.96	8.35	5.78	7.65	7.49	7.21
Malic acid	4.23	3.37	3.62	3.65	3.17	3.42	3.14	3.94	3.08	3.49	3.14	3.01	3.23	2.95	3.33	3.20	3.17

Numbers in row are operating conditions of the central composite design (Table 2.2).

Table 2.5. Experimentally obtained retention times for each operating condition of the full factorial design using amine column.

Analyte	1	2	3	4	5	6	7	8	9	10	11
Acetone	1.88	1.98	1.88	2.06	1.89	2.04	1.98	2.07	1.93	1.95	1.95
Toluene	1.52	1.73	1.52	1.80	1.55	1.75	1.61	1.76	1.61	1.62	1.62
2'-Deoxyadenosine	2.58	4.96	2.63	5.24	2.64	5.83	2.77	6.01	3.32	3.33	3.37
2'-Deoxycytidine	3.08	11.06	3.39	11.45	3.19	16.53	3.52	16.53	5.00	4.91	4.95
2',3'-Dideoxyadenosine	2.37	3.85	2.42	4.08	2.46	4.15	2.57	4.24	2.86	2.92	2.96
2'-Deoxyguanosine	3.25	11.91	3.32	12.12	3.28	18.52	3.43	17.26	4.95	4.82	4.83
3'-Deoxyguanosine	3.16	11.14	3.22	11.32	3.23	17.31	3.33	15.81	4.70	4.58	4.59
5'-Methyluridine	2.63	5.18	2.66	5.19	2.64	7.68	2.78	7.07	3.34	3.28	3.26
8-Hydroxy-2'-deoxyguanosine	3.07	10.44	3.21	10.49	3.07	18.25	3.30	16.09	4.63	4.48	4.48
Adenosine	2.77	6.18	2.83	6.28	2.83	8.79	2.99	8.20	3.72	3.70	3.72
Cytidine	3.51	15.38	3.81	15.22	3.70	32.87	3.98	25.89	6.02	5.82	5.82
Guanosine	3.60	16.29	3.68	15.85	4.12	32.85	3.83	26.47	5.86	5.61	5.61
Inosine	3.31	10.89	3.49	11.12	3.34	21.27	3.62	18.01	5.10	4.93	4.91
Thymidine	2.38	3.87	2.41	3.86	2.50	5.05	2.53	4.68	2.86	2.85	2.84
Uridine	2.82	5.95	2.86	5.85	3.01	10.47	3.00	8.91	3.70	3.62	3.59
Adrenaline	1.86	10.82	7.92	-	2.55	22.18	-	-	6.62	-	5.87

Noradrenaline	1.97	14.71	11.32	-	2.82	35.90	-	-	8.95	-	8.27
3'-Methoxytyramine	1.68	6.17	1.78	7.12	2.46	9.01	2.26	9.35	2.82	2.84	2.96
Isoproterenol	1.71	7.13	6.30	-	4.47	11.51	-	-	6.08	-	6.39
Fenotrole	1.55	5.11	1.61	5.29	4.42	6.31	1.92	6.29	2.35	2.33	2.38
Terbutaline	1.63	5.71	1.72	6.31	4.54	7.69	2.09	7.92	2.64	2.64	2.71
Salbutamol	1.68	6.51	1.77	7.43	4.60	9.30	2.29	9.66	2.86	2.87	2.97
Ritudrine	1.48	3.77	1.53	4.03	3.56	4.29	1.80	4.43	2.09	2.10	2.14
Metaproterenol	1.69	6.63	1.79	7.47	4.08	9.68	2.20	9.84	2.86	2.85	2.93
Synephrine	1.73	6.84	1.88	8.33	4.09	10.47	2.33	11.16	3.06	3.08	3.21
Dopamine	1.79	9.23	3.89	-	4.35	17.03	6.23	33.35	5.13	4.31	4.52
N-methylephrine	1.54	3.80	1.63	4.44	3.57	4.16	1.99	-	2.27	2.33	2.43
Norphenylephrine	1.79	8.81	1.92	9.63	4.20	14.60	2.37	14.27	3.32	3.27	3.37
Phenylephrine	1.72	6.73	1.85	7.96	3.87	9.68	2.28	10.45	3.00	3.01	3.12
Tyramine	1.68	6.18	1.77	6.99	3.31	8.99	2.17	9.62	2.81	2.80	2.90
Normetanephrine	1.79	8.79	1.94	9.99	3.10	12.34	2.44	14.80	3.42	3.38	3.52
Octopamine	1.79	8.93	1.92	9.97	2.96	13.00	2.42	15.40	3.41	3.35	3.47
Methoxamine	1.57	4.66	1.63	5.06	2.58	5.05	1.96	5.92	2.38	2.39	2.46
Isoxuprine	1.41	2.87	1.46	2.65	2.17	2.85	1.66	2.79	1.84	1.85	1.87
Ceftiofur	5.18	12.02	5.85	13.28	4.87	7.46	3.88	9.11	5.42	5.07	5.02
Tyrosine	4.51	45.31	5.00	52.85	6.07	63.01	5.12	77.53	9.98	9.12	9.07
Pentoxyfylline	1.81	2.19	1.84	2.25	2.69	2.09	1.94	2.15	1.97	2.00	2.02
Guanine	3.23	10.41	3.35	10.46	5.45	13.67	3.44	11.07	4.88	4.66	4.61
Xanthine	2.69	5.50	3.45	6.59	5.05	7.40	3.36	7.00	4.05	3.87	3.81

Caffeine	1.98	2.35	2.00	2.42	3.47	2.30	2.10	2.32	2.15	2.18	2.19
Theophylline	2.11	2.78	2.20	2.92	3.26	2.84	2.29	2.82	2.42	2.42	2.42
Theobromine	2.22	2.93	2.24	2.98	3.34	3.08	2.35	2.94	2.53	2.53	2.53
Diphylline	2.37	3.75	2.40	3.82	3.73	4.09	2.50	3.70	2.89	2.86	2.85
7-Hydroxyethyltheophylline	2.15	2.84	2.18	2.91	3.52	2.89	2.27	2.82	2.45	2.45	2.46
1-Methyluric acid	3.12	8.38	13.58	28.34	4.24	10.71	8.80	23.97	13.61	11.97	11.39
1-Methylguanine	2.78	6.15	2.87	6.31	3.76	7.16	2.97	6.32	3.77	3.69	3.69
9-Methylguanine	2.95	7.22	2.98	7.17	4.20	8.42	3.08	7.23	4.02	3.91	3.90
Uric acid	4.35	18.40	20.32	63.57	6.02	29.53	12.01	56.88	23.96	20.01	18.62
3,7-Dimethyluric acid	2.75	5.61	11.00	19.06	4.55	7.06	7.58	15.66	10.37	9.49	9.20
7-Methylxanthine	2.44	3.96	2.55	4.08	4.30	4.55	2.61	4.17	3.01	2.97	2.96
Hypoxanthine	2.92	6.56	3.13	6.96	4.60	8.48	3.18	7.26	4.12	3.98	3.94
Proxyphylline	2.04	2.61	2.07	2.68	2.90	2.60	2.15	2.57	2.29	2.30	2.31
1,7-Dimethyluric acid	2.76	5.67	11.13	18.44	3.33	6.27	7.44	14.44	10.24	9.31	8.99
1,3-Dimethyluric acid	2.49	4.75	7.87	13.29	3.17	5.55	5.92	10.96	7.18	6.71	6.60
1,3,7-Trimethyluric acid	2.28	3.20	2.31	3.30	3.01	3.26	2.41	3.18	2.62	2.65	2.67
4-Nitrophenyl-B-D-glycopyranoside	2.06	4.02	2.08	4.00	2.24	4.23	2.15	-	2.58	2.50	2.47
Acyclovir	3.20	10.31	3.25	10.18	3.56	13.63	3.33	10.80	4.82	4.61	4.58
2'-Deoxyuridine	2.53	4.35	2.56	4.36	3.00	5.45	2.63	4.57	3.19	3.10	3.07
3'-Deoxythymidine	2.12	2.86	2.14	2.93	2.72	2.99	2.22	2.86	2.41	2.40	2.40
2'-Deoxyinosine	3.01	8.14	3.14	8.40	4.18	11.06	3.21	8.89	4.47	4.28	4.24
Satolol	1.58	4.59	1.65	5.19	2.67	5.02	1.95	5.08	2.70	2.63	2.64

7-Methylguanosine	2.12	20.51	3.02	29.28	4.32	36.27	3.75	32.93	6.46	-	6.51
Atenolol	1.71	7.67	1.81	9.03	3.91	9.32	2.31	9.01	3.09	3.12	3.26
Vadarabine	2.86	7.34	2.94	7.49	5.98	9.63	3.03	7.76	4.11	3.98	3.97
Tryptophan	3.43	25.78	3.63	28.64	7.30	27.47	3.72	25.47	6.35	5.95	6.06
BTMA	1.54	3.59	1.62	4.30	3.91	3.96	1.99	4.70	2.25	2.35	2.47
PTMA	1.60	4.01	1.70	4.88	4.06	4.68	2.13	5.11	2.43	2.56	2.72
Labetalol	1.45	3.44	1.52	3.74	3.35	3.36	1.75	3.82	-	-	-
Nadolol	1.66	7.26	1.76	8.26	3.82	8.30	-	8.14	2.88	2.89	2.99
Propranolol	1.43	3.19	1.48	3.49	2.89	-	1.70	-	1.94	1.94	
Adenine	2.64	5.24	2.79	5.47	4.34	5.89	2.89	5.72	3.48	3.47	3.48
Uracil	2.44	3.49	2.48	3.57	3.62	4.20	2.56	3.68	2.87	2.82	2.79
Thymine	2.27	3.14	2.30	3.19	3.28	3.53	2.37	3.24	2.62	2.58	2.56
Cytosine	2.86	9.55	3.41	9.65	4.44	12.78	3.52	10.24	4.89	4.73	4.70
Pindolol	1.47	3.46	1.52	3.83	2.43	-	1.78	3.71	-	-	-
Alprenolol	1.42	3.12	1.47	3.40	2.42	-	1.69	-	1.89	1.91	1.95
Salicylic acid	5.23	4.75	5.72	4.87	5.86	3.74	3.73	3.71	4.16	3.92	3.77
5-Methylsalicylic acid	4.27	4.32	5.13	4.63	3.65	3.41	3.46	3.56	3.81	3.62	3.51
4-Hydroxybenzoic acid	2.06	2.83	12.83	12.62	2.17	2.98	7.92	9.82	8.76	7.87	7.47
3-Hydroxybenzoic acid	2.54	3.85	12.85	16.93	2.58	4.16	7.61	12.32	10.39	9.07	8.53
2,3-Dihydroxybenzoic acid	7.45	7.72	10.72	7.89	-	-	5.69	6.08	6.32	5.66	5.26
2,4-Dihydroxybenzoic acid	4.83	6.30	7.88	8.20	4.47	5.62	4.94	6.24	6.06	5.51	5.19
2,5-Dihydroxybenzoic acid	6.70	7.44	7.13	7.54	5.79	6.19	4.63	5.90	5.64	5.15	4.87
3,4-Dihydroxybenzoic acid	2.42	4.35	-	-	2.79	5.64	-	-	6.54	-	-

3,5-Dihydroxybenzoic acid	3.00	5.61	15.28	26.92	3.24	6.87	8.82	20.76	13.65	11.64	10.76
Benzoic acid	2.26	2.90	10.44	11.22	2.72	2.91	6.61	7.91	7.82	7.21	6.75
3-Amino-4-hydroxybenzoic acid	2.66	3.54	16.64	18.60	3.37	3.82	10.09	13.58	12.91	11.65	10.96
Gallic acid	3.11	11.75	-	-	-	-	-	-	-	-	-
4-Aminosalicylic acid	3.65	2.42	11.49	9.40	5.78	4.54	5.82	6.82	7.04	6.51	6.02
4-Aminobenzoic acid	1.94	4.67	9.14	8.02	3.30	2.47	7.63	6.56	6.67	6.29	6.15
Phthalic acid	28.57	12.65	-	-	23.29	11.77	-	76.55	118.59	118.59	77.34
Isophthalic acid	26.02	45.91	-	-	19.85	60.27	-	-	-	-	216.40
3-Aminobenzoic acid	2.36	3.54	19.73	20.42	3.04	4.02	12.05	16.59	13.64	12.20	11.40
Vanillic acid	2.14	2.95	13.62	13.81	2.81	3.08	8.37	10.37	9.37	8.59	8.17
Syringic acid	2.30	3.32	14.42	16.38	3.05	3.46	8.90	11.82	10.47	9.62	9.17
2-Methoxybenzoic acid	2.59	3.31	15.12	16.76	3.36	3.41	9.25	11.86	11.59	10.55	9.92
Terephthalic acid	27.80	43.25	-	-	21.23	60.88	-	-	-	-	-
Caffeic acid	2.17	3.70	-	-	3.09	4.12	-	-	-	-	-
P-Toluic acid	1.92	2.48	8.32	8.91	2.33	2.43	5.52	6.38	5.93	5.64	5.47
3,5-Dinitrosalicylic acid	2.50	1.85	2.72	1.84	2.15	1.72	2.18	1.75	1.93	1.94	1.92
Nicotinic acid	11.59	19.61	20.48	39.71	8.63	19.98	12.43	26.99	21.19	19.04	17.87
4-Hydroxybenzenesulfonic acid	9.75	9.99	8.43	8.88	5.84	9.28	5.27	7.51	6.75	6.06	5.61
4-Aminophenylacetic acid	2.87	3.85	21.76	24.13	2.73	4.16	13.30	17.63	16.68	15.30	14.46
Acetylsalicylic acid	3.25	4.37	10.90	13.90	2.92	4.24	6.84	9.32	8.99	8.31	7.89
Phenylacetic acid	2.17	2.78	11.17	12.40	2.14	2.75	7.15	8.74	8.25	7.78	7.48
1,2-benzenedisulfonic acid	-	-	-	-	-	-	-	-	-	-	-
P-Toluenesulfonic acid	6.86	5.56	5.48	4.96	4.45	4.43	3.73	-	4.25	4.02	3.83

2-Sulfobenzoic acid	-	204.10	-	-	174.15	-	-	-	-	-	-
Tropic acid	3.01	4.88	13.73	21.66	2.80	5.09	8.41	-	12.09	10.90	10.27
2-Phenylethylamine	1.58	4.59	1.67	5.12	1.87	5.18	2.01	5.61	2.39	2.43	2.49
Phenylalanine	3.84	27.46	4.10	29.47	3.94	-	4.19	28.39	7.19	6.75	6.78
Mandelic acid	8.18	11.72	12.66	16.60	6.36	-	7.79	11.77	10.79	9.62	8.88
5-Sulfosalicylic acid	-	104.06	-	-	-	-	-	-	61.83	48.61	40.67
Malic acid	10.57	5.91	72.42	16.93	5.95	5.82	27.26	14.30	24.02	19.27	15.92

Numbers in row are operating conditions of the full factorial design (Table 2.3).

Table 2.6. Experimentally obtained retention times for each operating condition of the full factorial design using zwitterionic column.

Analyte	1	2	3	4	5	6	7	8	9	10	11
Acetone	1.91	1.92	1.88	1.91	1.89	1.88	1.87	1.89	1.89	1.89	1.85
Toluene	1.60	1.69	1.56	1.68	1.57	1.63	1.55	1.64	1.59	1.59	1.56
2'-Deoxyadenosine	2.58	4.93	2.52	5.01	2.57	5.33	2.55	5.35	3.23	3.22	3.16
2'-Deoxycytidine	3.21	10.59	3.00	10.82	3.23	13.28	3.04	13.23	4.37	4.47	4.49
2',3'-Dideoxyadenosine	2.46	3.71	2.35	3.70	2.43	3.88	2.39	3.87	2.88	2.81	2.73
2'-Deoxyguanosine	2.95	11.69	2.98	12.21	2.95	14.01	2.99	14.12	4.28	4.45	4.51
3'-Deoxyguanosine	2.87	10.79	2.89	11.21	2.87	12.79	2.90	12.90	4.06	4.21	4.26
5'-Methyluridine	2.41	4.73	2.43	4.88	2.42	5.65	2.44	5.64	2.93	3.02	3.03
8-Hydroxy-2'-deoxyguanosine	2.78	10.61	2.83	11.19	2.79	12.92	2.84	13.11	3.94	4.13	4.21
Adenosine	2.68	6.05	2.65	6.23	2.68	6.98	2.67	7.05	3.47	3.51	3.49
Cytidine	3.39	13.99	3.22	14.64	3.43	20.20	3.25	20.24	4.89	5.09	5.18
Guanosine	3.12	15.81	3.20	16.82	3.14	20.79	3.20	21.05	4.79	5.06	5.23

Inosine	2.86	9.68	2.93	10.43	2.89	13.38	2.95	13.69	4.13	4.29	4.36
Thymidine	2.28	3.63	2.27	3.69	2.27	3.93	2.27	3.96	2.63	2.67	2.66
Uridine	2.55	5.58	2.59	5.81	2.57	7.09	2.60	7.10	3.21	3.33	3.38
Adrenaline	3.44	17.30	6.00	-	3.36	19.25	5.07	-	-	6.75	7.03
Noradrenaline	3.70	24.97	6.81	-	3.62	-	6.04	-	-	8.14	8.36
3'-Methoxytyramine	2.88	9.30	4.06	11.24	2.80	9.04	3.25	10.22	4.65	4.39	4.39
Isoproterenol	2.92	10.57	5.99	-	2.84	-	5.47	-	-	5.42	5.28
Fenotrole	2.42	9.44	3.11	9.71	2.33	6.56	2.52	6.87	3.33	3.30	3.37
Terbutaline	2.70	9.57	3.73	10.83	2.62	7.75	2.95	8.59	4.07	3.94	3.99
Salbutamol	2.80	9.47	3.89	11.22	2.73	8.92	3.11	9.99	4.48	4.23	4.24
Ritudrine	2.23	5.72	2.76	6.13	2.14	4.39	2.29	4.63	2.88	2.79	2.80
Metaproterenol	2.87	11.52	4.12	13.29	2.80	9.71	3.20	10.85	4.57	4.45	4.55
Synephrine	3.00	10.59	4.46	13.20	2.93	10.42	3.46	12.14	5.11	4.84	4.90
Dopamine	3.21	15.02	5.23	-	3.15	15.89	4.22	-	-	5.75	5.91
N-methylephrine	2.45	4.73	3.29	5.53	2.38	4.53	2.70	4.93	3.55	3.22	3.12
Norphenylephrine	3.16	15.24	4.76	16.80	3.09	13.80	3.57	15.02	5.38	5.27	5.45
Phenylephrine	2.95	10.41	4.36	12.63	2.89	9.81	3.38	11.19	4.94	4.67	4.73
Tyramine	2.83	9.55	4.08	11.44	2.77	9.10	3.18	10.23	4.50	4.32	4.37
Normetanephrine	3.23	14.14	4.98	16.89	3.16	14.11	3.73	15.68	5.74	5.51	5.65
Octopamine	3.21	14.85	4.97	17.55	3.14	14.62	3.67	16.25	5.58	5.47	5.66
Methoxamine	2.46	6.46	3.24	6.95	2.38	5.53	2.61	5.86	3.46	3.27	3.23
Isoxuprine	2.01	3.65	2.27	3.14	1.93	3.05	1.97	2.66	-	2.24	2.18
Ceftiofur	2.01	5.60	1.88	6.32	1.95	5.49	1.95	5.87	2.38	2.48	2.50

Tyrosine	3.64	37.59	3.93	45.22	3.61	52.33	3.86	60.52	7.10	7.25	7.46
Pentoxyfylline	1.86	2.06	1.89	2.06	1.84	1.99	1.82	2.00	1.95	1.92	1.87
Guanine	3.04	10.87	3.20	11.55	3.04	12.56	3.09	13.19	4.27	4.43	4.51
Xanthine	2.46	5.13	2.65	5.96	2.49	6.29	2.58	7.45	3.16	3.24	3.27
Caffeine	2.00	2.20	2.04	2.21	1.98	2.18	1.97	2.19	2.11	2.08	2.03
Theophylline	2.09	2.60	2.16	2.66	2.07	2.63	2.08	2.68	2.27	2.25	2.22
Theobromine	2.20	2.74	2.26	2.76	2.19	2.86	2.19	2.88	2.43	2.43	2.38
Diphylline	2.28	3.36	2.35	3.41	2.27	3.66	2.28	3.71	2.66	2.66	2.61
7-Hydroxyethyltheophylline	2.13	2.64	2.19	2.66	2.11	2.68	2.11	2.70	2.33	2.28	2.25
1-Methyluric acid	2.46	6.32	2.92	16.52	2.47	8.22	3.09	23.50	4.66	4.86	5.05
1-Methylguanine	2.70	5.92	2.82	6.17	2.69	6.60	2.71	6.85	3.46	3.48	3.45
9-Methylguanine	2.84	7.06	2.96	7.23	2.83	7.82	2.84	8.03	3.70	3.73	3.72
Uric acid	2.80	12.21	3.31	37.50	2.83	21.02	3.59	64.44	6.38	6.97	7.36
3,7-Dimethyluric acid	2.35	4.39	2.82	10.84	2.35	5.68	2.93	15.20	4.21	4.23	4.30
7-Methylxanthine	2.33	3.65	2.45	3.82	2.33	4.09	2.35	4.22	2.74	2.75	2.72
Hypoxanthine	2.68	6.08	2.87	6.61	2.69	7.36	2.76	7.92	3.52	3.58	3.58
Proxyphylline	2.04	2.42	2.11	2.46	2.02	2.42	2.02	2.43	2.20	2.17	2.13
1,7-Dimethyluric acid	2.30	4.33	2.70	10.12	2.29	5.00	2.82	13.16	3.97	4.02	4.10
1,3-Dimethyluric acid	2.26	3.96	2.61	7.92	2.26	4.60	2.64	9.97	3.49	3.47	3.51
1,3,7-Trimethyluric acid	2.25	2.89	2.32	2.95	2.23	3.04	2.24	3.06	2.54	2.49	2.43
4-Nitrophenyl-B-D-glycopyranoside	2.02	3.65	2.11	3.75	2.01	3.67	1.99	3.76	2.27	2.32	-
Acyclovir	2.91	9.70	3.08	10.25	2.93	11.92	2.96	12.62	4.15	4.24	4.25

2'-Deoxyuridine	2.37	4.15	2.52	4.29	2.39	4.73	2.41	4.93	2.84	2.92	2.91
3'-Deoxythymidine	2.08	2.65	2.16	2.70	2.07	2.71	2.06	2.75	2.26	2.27	2.21
2'-Deoxyinosine	2.70	7.32	2.88	7.96	2.73	9.26	2.77	10.02	3.73	3.82	3.82
Satolol	2.50	6.72	3.44	7.45	2.43	5.65	2.69	6.05	3.62	3.41	3.37
7-Methylguanosine	4.06	32.69	6.07	43.71	4.03	38.45	4.82	50.77	9.43	9.01	9.15
Atenolol	2.96	11.62	4.31	14.13	2.89	10.92	3.35	12.09	5.28	4.83	4.71
Vadarabine	2.79	7.63	2.90	8.17	2.80	8.83	2.78	9.30	3.75	3.84	3.83
Tryptophan	3.23	23.75	3.48	27.60	3.17	25.79	3.30	29.17	5.32	5.42	5.54
BTMA	2.61	4.94	3.74	5.80	2.58	5.10	3.02	5.39	4.13	3.55	3.30
PTMA	2.81	5.61	4.23	6.66	2.80	6.26	3.41	6.78	4.88	4.06	3.74
Labetalol	2.20	-	2.64	5.25	2.14	-	2.23	-	2.70	2.63	2.60
Nadolol	2.76	10.34	3.91	11.86	2.69	9.17	3.03	10.01	4.53	4.22	4.14
Propranolol	2.20	4.14	2.62	4.52	2.14	-	2.23	-	2.66	2.60	2.57
Adenine	2.79	5.48	2.85	5.84	2.79	5.82	2.73	6.04	3.44	3.44	3.37
Uracil	2.31	3.35	2.45	3.58	2.32	3.73	2.34	3.91	2.62	2.68	2.67
Thymine	2.18	2.98	2.30	3.16	2.19	3.16	2.20	3.27	2.41	2.46	2.45
Cytosine	3.37	9.39	3.23	9.78	3.39	11.54	3.11	12.24	4.29	4.37	4.34
Pindolol	2.29	5.22	2.49	4.08	2.25	-	2.41	4.43	3.02	2.90	2.86
Alprenolol	2.17	3.75	2.50	4.11	2.06	-	2.19	-	2.54	2.57	2.55
Salicylic acid	2.07	3.14	1.97	3.44	2.00	3.06	2.00	3.22	2.16	2.30	2.38
5-Methylsalicylic acid	1.94	2.88	1.89	3.29	1.89	2.81	1.93	2.99	2.07	2.19	2.19
4-Hydroxybenzoic acid	1.80	2.39	2.82	7.90	1.78	2.59	2.77	8.48	3.35	3.41	3.62
3-Hydroxybenzoic acid	1.84	2.80	2.70	10.31	1.82	3.42	2.75	10.42	3.53	3.78	3.99

2,3-Dihydroxybenzoic acid	2.43	4.76	2.24	5.31	2.28	4.68	2.37	4.96	2.64	2.86	2.99
2,4-Dihydroxybenzoic acid	2.10	4.39	2.25	6.33	2.04	4.60	2.32	5.64	2.69	2.97	3.10
2,5-Dihydroxybenzoic acid	2.33	4.99	2.16	5.68	2.23	4.92	2.24	5.26	2.56	2.82	2.91
3,4-Dihydroxybenzoic acid	1.92	3.17	3.52	25.37	1.90	4.13	3.98	-	-	4.97	5.77
3,5-Dihydroxybenzoic acid	1.96	4.01	2.99	19.42	1.93	5.48	3.06	19.48	4.27	4.78	5.21
Benzoic acid	1.75	2.20	2.45	6.17	1.72	2.46	2.50	6.26	3.00	3.10	3.16
3-Amino-4-hydroxybenzoic acid	1.97	2.93	3.60	13.59	1.95	3.26	3.53	11.92	4.33	4.71	5.14
Gallic acid	2.10	5.29	-	-	2.06	-	6.60	-	-	-	-
4-Aminobenzoic acid	1.82	2.21	2.92	5.56	1.80	2.26	2.78	5.93	3.10	3.07	3.22
4-Aminosalicylic acid	1.99	3.50	2.57	7.64	1.94	3.99	2.64	6.64	3.15	3.52	3.71
Phthalic acid	2.15	3.21	2.88	13.25	2.09	4.91	3.16	-	4.23	4.25	4.44
Isophthalic acid	2.10	8.74	5.96	218.91	2.04	28.86	6.40	-	16.34	19.40	22.16
3-Aminobenzoic acid	1.91	2.65	3.19	11.20	1.89	3.44	3.25	15.68	4.64	4.52	4.74
Vanillic acid	1.82	2.40	2.87	8.47	1.79	2.67	2.88	8.96	3.58	3.65	3.82
Syringic acid	1.86	2.58	2.96	9.89	1.83	2.95	3.02	10.23	3.94	4.03	4.19
2-Methoxybenzoic acid	1.79	2.23	2.67	7.49	1.76	2.62	2.74	8.57	3.48	3.54	3.59
Terephthalic acid	2.16	9.45	6.89	276.85	2.10	30.65	7.47	-	20.73	24.54	28.54
Caffeic acid	1.85	2.87	3.16	18.78	1.82	3.35	3.47	-	-	4.23	4.82
P-Toluic acid	1.69	2.03	2.26	4.89	1.67	2.14	2.28	4.90	2.65	2.68	2.70
3,5-Dinitrosalicylic acid	1.58	1.58	1.51	1.64	1.57	1.54	1.51	1.61	1.50	1.55	1.52
Nicotinic acid	3.02	8.58	3.34	19.89	2.90	13.91	3.63	23.69	5.93	6.35	6.30
4-Hydroxybenzenesulfonic acid	2.56	6.42	2.18	6.45	2.49	6.81	2.29	7.13	2.67	3.01	3.14

4-Aminophenylacetic acid	2.06	2.76	3.67	10.05	2.04	3.39	3.69	14.63	5.11	5.12	5.32
Acetylsalicylic acid	1.82	2.63	2.38	6.29	1.78	3.32	2.47	7.09	3.10	3.25	3.25
Phenylacetic acid	1.73	2.12	2.51	5.74	1.71	2.32	2.54	6.52	3.11	3.25	3.13
1,2-benzenedisulfonic acid	6.04	50.10	3.26	47.32	4.58	76.16	3.56	-	6.23	7.69	8.53
P-Toluenesulfonic acid	2.19	3.47	1.90	3.26	2.13	3.43	1.97	3.51	2.13	2.31	-
2-Sulfobenzoic acid	2.93	16.38	4.19	79.31	2.80	56.66	4.53	-	9.48	11.22	12.19
Tropic acid	1.89	2.90	2.58	9.65	1.86	3.78	2.69	11.37	3.62	3.80	3.83
2-Phenylethylamine	2.53	6.54	3.52	7.63	2.47	5.76	2.75	6.52	3.65	3.50	3.40
Phenylalanine	3.17	20.01	3.36	24.38	3.14	26.21	3.29	29.92	-	5.41	5.32
Mandelic acid	2.30	5.07	2.41	8.07	2.18	-	2.58	-	-	3.52	3.54
5-Sulfosalicylic acid	4.71	30.28	2.68	30.61	3.81	33.50	3.01	34.31	4.61	5.82	6.50
Malic acid	2.33	3.24	2.13	4.99	2.27	3.79	2.28	9.22	2.60	2.77	2.78

Numbers in row are operating conditions of the full factorial design (Table 2.3).

Table 2.7. Experimentally obtained retention times for each operating condition of the full factorial design using bare silica column.

Analyte	1	2	3	4	5	6	7	8	9	10	11
Acetone	1.12	1.14	1.12	1.13	1.11	1.12	1.14	1.14	1.11	1.20	1.12
Toluene	0.98	1.02	0.98	1.02	0.98	1.00	0.99	1.00	0.98	1.07	0.99
2'-Deoxyadenosine	1.41	2.57	1.42	2.58	1.43	2.70	1.45	2.75	1.69	1.80	1.71
2'-Deoxycytidine	1.54	4.32	1.52	4.50	1.55	5.86	1.56	6.89	2.00	2.12	2.04
2',3'-Dideoxyadenosine	1.48	2.34	1.46	2.29	1.48	2.24	1.47	2.11	1.68	1.79	1.69
2'-Deoxyguanosine	1.38	3.82	1.44	4.06	1.42	5.46	1.48	6.52	1.85	1.95	1.88

3'-Deoxyguanosine	1.37	3.54	1.41	3.75	1.39	5.05	1.45	5.96	1.78	1.88	1.81
5'-Methyluridine	1.22	1.87	1.25	1.94	-	2.44	1.29	2.80	1.40	1.50	1.42
8-Hydroxy-2'-deoxyguanosine	1.31	3.11	1.36	3.36	1.34	4.78	1.40	6.02	1.68	1.81	1.72
Adenosine	1.38	2.69	1.41	2.76	1.41	3.18	1.45	3.52	1.70	1.85	1.73
Cytidine	1.53	4.83	1.53	5.19	1.55	7.99	1.59	10.86	2.07	2.23	2.12
Guanosine	1.38	4.22	1.45	4.62	1.42	7.26	1.50	9.65	1.90	2.04	1.94
Inosine	1.35	3.31	1.42	3.62	1.39	5.22	1.47	7.13	1.78	1.93	1.82
Thymidine	1.20	1.67	1.23	1.70	1.22	1.89	1.26	2.04	1.34	1.50	1.36
Uridine	1.24	2.03	1.29	2.13	1.28	2.87	1.33	3.60	1.47	1.63	1.49
Adrenaline	1.68	6.73	-	-	1.63	7.22	5.77	-	3.15	3.93	3.66
Noradrenaline	1.67	7.30	-	-	1.62	8.96	5.10	-	3.13	4.19	3.93
3'-Methoxytyramine	1.58	4.86	2.30	5.59	1.54	4.18	1.98	4.63	2.34	2.70	2.39
Isoproterenol	1.55	4.63	-	-	1.51	4.37	4.89	-	2.84	3.71	3.37
Fenotrole	1.30	2.72	1.58	2.83	1.27	2.50	1.40	2.47	1.59	1.80	1.61
Terbutaline	1.44	3.53	1.96	4.00	1.42	3.26	1.68	3.53	1.96	2.24	1.98
Salbutamol	1.57	4.59	2.25	5.35	1.53	4.17	1.90	4.65	2.30	2.63	2.33
Ritudrine	1.28	2.47	1.55	2.56	1.25	2.04	1.38	1.99	1.53	1.76	1.54
Metaproterenol	1.49	4.10	2.08	4.74	1.47	3.93	1.78	4.44	2.13	2.43	2.15
Synephrine	1.60	5.20	2.43	6.37	1.58	4.67	2.04	5.53	2.52	2.94	2.58
Dopamine	1.59	5.62	-	-	1.58	5.79	3.28	12.20	2.63	3.24	2.97
N-methylephrine	1.58	3.90	2.40	4.48	1.55	2.82	2.01	2.74	2.28	2.73	2.31
Norphenylephrine	1.57	5.26	2.25	6.08	1.55	5.31	1.90	6.22	2.38	2.80	2.42
Phenylephrine	1.58	4.92	2.36	5.96	1.55	4.39	1.99	5.05	2.43	2.88	2.48

Tyramine	1.53	4.54	2.18	5.19	1.49	4.03	1.84	4.51	2.21	2.61	2.26
Normetanephrine	1.63	5.85	2.45	6.88	1.59	5.71	2.11	6.83	2.59	3.04	2.66
Octopamine	1.59	5.52	2.32	6.47	1.55	5.62	1.99	6.82	2.45	2.86	2.52
Methoxamine	1.46	3.61	1.96	3.90	1.42	2.89	1.67	2.82	1.93	2.29	1.96
Isoxuprine	1.25	2.11	1.40	1.75	1.20	1.68	1.27	1.38	1.37	1.61	1.38
Ceftiofur	1.06	2.23	0.99	2.37	1.09	2.41	1.08	2.59	1.22	1.38	1.22
Tyrosine	1.61	13.03	1.76	16.70	1.66	18.83	1.96	28.23	2.86	3.31	2.97
Pentoxyfylline	1.18	1.32	1.17	1.31	1.18	1.24	1.18	1.18	1.20	1.40	1.21
Guanine	1.43	3.64	1.47	3.83	1.47	4.90	1.52	5.81	1.86	2.08	1.88
Xanthine	1.26	2.04	1.30	2.29	1.28	2.69	1.35	3.59	1.49	1.69	1.51
Caffeine	1.23	1.39	1.22	1.37	1.23	1.33	1.24	1.27	1.27	1.47	1.28
Theophylline	1.21	1.46	1.22	1.48	1.23	1.48	1.24	1.47	1.28	1.49	1.30
Theobromine	1.28	1.60	1.29	1.59	1.31	1.63	1.32	1.60	1.38	1.59	1.39
Diphylline	1.28	1.87	1.31	1.90	1.33	1.95	1.34	1.97	1.46	1.68	1.47
7-Hydroxyethyltheophylline	1.25	1.54	1.26	1.54	1.28	1.55	1.28	1.50	1.34	1.55	1.36
1-Methyluric acid	1.23	2.37	1.25	4.99	1.28	3.11	1.39	10.56	1.82	2.06	1.76
1-Methylguanine	1.43	2.86	1.45	2.92	1.48	3.17	1.48	3.36	1.74	1.99	1.76
9-Methylguanine	1.46	3.10	1.48	3.15	1.57	3.64	1.52	3.87	1.80	2.05	1.82
Uric acid	1.28	3.83	1.29	9.13	1.34	6.48	1.47	28.11	2.18	2.44	2.25
3,7-Dimethyluric acid	1.25	2.18	1.29	4.35	1.30	2.52	1.43	7.10	1.83	2.08	1.87
7-Methylxanthine	1.28	1.84	1.30	1.88	1.32	2.08	1.33	2.20	1.44	1.66	1.46
Hypoxanthine	1.38	2.67	1.42	2.84	1.43	3.33	1.46	3.87	1.70	1.95	1.72
Proxyphylline	1.23	1.45	1.22	1.45	1.25	1.43	1.24	1.38	1.29	1.51	1.30

1,7-Dimethyluric acid	1.21	1.99	1.24	3.89	1.26	2.22	1.37	6.07	1.73	1.95	1.76
1,3-Dimethyluric acid	1.22	1.92	1.25	3.36	1.27	2.14	1.36	4.71	1.62	1.86	1.65
1,3,7-Trimethyluric acid	1.37	1.90	1.38	1.89	1.42	1.84	1.40	1.78	1.51	1.77	1.52
4-Nitrophenyl-B-D-glycopyranoside	1.10	1.50	1.12	1.56	1.14	1.73	1.14	1.79	1.20	1.38	1.21
Acyclovir	1.44	3.84	1.49	4.02	1.52	5.19	1.54	6.00	1.92	2.22	1.94
2'-Deoxyuridine	1.23	1.80	1.26	1.85	1.30	2.17	1.30	2.41	1.41	1.65	1.43
3'-Deoxythymidine	1.19	1.44	1.20	1.45	1.23	1.51	1.22	1.51	1.27	1.50	1.28
2'-Deoxyinosine	1.36	3.05	1.42	3.25	1.43	4.07	1.46	4.84	1.76	2.04	1.79
Satolol	1.48	3.77	2.04	4.27	1.44	3.02	1.75	3.05	2.05	2.45	2.07
7-Methylguanosine	2.06	13.44	3.09	18.27	2.12	15.84	2.71	24.52	4.28	5.18	4.43
Atenolol	1.79	7.35	2.76	8.59	1.83	5.92	2.31	6.25	3.03	3.72	3.08
Vadarabine	1.40	3.00	1.43	3.11	1.49	3.75	1.47	4.17	1.77	2.05	1.79
Tryptophan	1.51	8.91	1.59	10.50	1.58	9.89	1.71	11.91	2.32	2.71	2.38
BTMA	2.08	5.70	3.48	6.13	2.10	3.87	2.77	3.68	3.17	3.85	3.23
PTMA	2.33	6.83	4.18	7.43	2.38	4.85	3.29	4.75	3.78	4.68	3.88
Labetalol	1.29	-	1.56	2.50	1.29	1.91	1.39	-	1.53	1.85	1.54
Nadolol	1.65	5.75	2.41	6.71	1.32	4.78	2.03	4.92	1.61	3.04	2.57
Propranolol	1.33	2.58	1.64	2.72	1.33	1.96	1.44	-	1.61	1.97	1.61
Adenine	1.52	2.80	1.51	2.81	1.59	2.90	1.54	2.92	1.80	2.14	1.81
Uracil	1.22	1.57	1.24	1.61	1.31	1.81	1.28	1.96	1.34	1.60	1.36
Thymine	1.19	1.46	1.21	1.49	1.27	1.62	1.23	1.69	1.28	1.52	1.29
Cytosine	1.68	4.19	1.59	4.09	1.74	5.25	1.64	5.91	2.06	2.42	2.08

Pindolol	1.37	2.80	1.74	2.97	1.40	2.15	1.53	2.10	1.56	2.08	1.72
Alprenolol	1.30	2.46	1.59	2.58	1.32	1.90	1.41	-	1.56	1.91	1.57
Salicylic acid	1.03	1.34	0.97	1.36	1.13	1.46	1.05	1.53	1.09	1.26	1.11
5-Methylsalicylic acid	1.03	1.31	0.96	1.34	1.09	1.40	1.04	1.46	1.08	1.29	1.09
4-Hydroxybenzoic acid	1.03	1.24	1.14	2.34	1.12	1.28	1.25	3.38	1.38	1.62	1.41
3-Hydroxybenzoic acid	1.04	1.40	1.11	2.77	1.16	1.47	1.22	4.14	1.43	1.69	1.46
2,3-Dihydroxybenzoic acid	1.09	1.55	1.01	1.63	1.08	1.87	1.15	2.11	1.18	1.40	1.20
2,4-Dihydroxybenzoic acid	1.05	1.54	1.00	1.66	1.13	1.77	1.09	2.26	1.18	1.41	1.19
2,5-Dihydroxybenzoic acid	1.06	1.56	0.99	1.57	1.13	1.91	1.07	2.17	1.16	1.34	1.17
3,4-Dihydroxybenzoic acid	1.06	1.44	1.22	7.44	1.14	1.61	1.77	-	1.63	2.11	1.90
3,5-Dihydroxybenzoic acid	1.05	1.60	1.11	3.39	1.08	1.85	1.23	7.09	1.48	1.73	1.52
Benzoic acid	1.03	1.30	1.11	2.39	1.10	1.28	1.22	2.77	1.39	1.63	1.41
3-Amino-4-hydroxybenzoic acid	1.08	-	1.23	3.43	1.17	1.46	1.38	4.91	1.69	1.97	1.73
Gallic acid	1.08	1.94	1.41	-	1.10	3.28	2.81	-	2.23	3.50	3.35
4-Aminobenzoic acid	1.05	1.53	1.04	1.93	1.13	1.60	1.15	2.52	1.29	1.60	1.30
4-Aminosalicylic acid	1.05	1.19	1.18	1.94	1.13	1.20	1.29	2.52	1.33	1.51	1.37
Phthalic acid	1.03	1.55	1.11	5.19	1.18	2.09	1.36	18.18	1.58	1.84	1.68
Isophthalic acid	1.08	5.29	1.36	39.85	1.14	6.87	1.72	130.25	3.83	4.00	4.00
3-Aminobenzoic acid	1.08	1.37	1.25	3.34	1.11	1.48	1.43	5.87	1.62	1.93	1.67
Vanillic acid	1.04	1.29	1.16	2.63	1.14	1.31	1.30	3.61	1.48	1.72	1.50
Syringic acid	1.06	1.40	1.20	3.25	1.13	1.41	1.36	4.16	1.60	1.87	1.63
2-Methoxybenzoic acid	1.05	1.37	1.19	3.08	1.22	1.36	1.33	3.97	1.59	1.86	1.61
Terephthalic acid	1.09	5.17	1.42	44.71	1.11	6.94	1.85	165.37	4.40	4.92	4.60

Caffeic acid	1.04	-	1.16	5.36	1.07	1.45	1.57	10.16	1.53	1.89	1.66
P-Toluic acid	1.02	1.21	1.09	2.10	0.99	1.18	1.19	2.22	1.31	1.57	1.33
3,5-Dinitrosalicylic acid	0.92	0.90	0.88	0.89	1.63	0.96	0.94	0.97	0.91	1.09	0.91
Nicotinic acid	1.48	5.66	1.43	8.25	1.17	5.84	1.67	11.45	2.63	3.03	2.63
4-Hydroxybenzenesulfonic acid	1.05	1.55	0.98	1.56	1.22	2.37	1.08	2.82	1.16	1.37	1.17
4-Aminophenylacetic acid	1.13	1.58	1.40	4.39	1.14	1.57	1.61	6.25	2.03	2.49	2.07
Acetylsalicylic acid	1.05	1.62	1.12	2.74	1.11	1.57	1.25	3.26	1.48	1.76	1.49
Phenylacetic acid	1.03	1.30	1.16	2.75	1.38	1.25	1.29	3.02	1.50	1.88	1.51
1,2-benzenedisulfonic acid	1.21	7.59	1.08	7.74	1.11	22.68	1.28	33.70	1.83	2.17	1.87
P-Toluenesulfonic acid	1.03	1.29	0.96	1.27	1.11	1.57	1.04	1.64	1.08	1.36	1.09
2-Sulfobenzoic acid	1.07	7.13	1.25	19.63	1.26	17.23	1.54	144.59	2.79	3.35	2.89
Tropic acid	1.07	1.70	1.16	3.68	1.14	1.70	1.31	5.11	1.62	2.04	1.64
2-Phenylethylamine	1.51	3.84	2.08	4.11	1.53	3.05	-	3.18	2.06	2.70	2.08
Phenylalanine	1.62	10.04	1.71	10.87	1.76	11.79	1.91	14.77	2.62	3.19	2.72
Mandelic acid	1.18	2.43	1.11	2.81	1.28	2.72	1.28	3.90	1.48	1.84	1.50
5-Sulfosalicylic acid	1.08	3.80	0.96	3.78	1.23	8.93	1.13	12.54	1.40	1.69	1.42
Malic acid	1.06	1.32	1.01	1.87	1.16	1.71	1.14	5.19	1.20	1.51	1.22

Numbers in row are operating conditions of the full factorial design (Table 2.3).

Table 2.8. Experimentally obtained retention times for each operating condition of the full factorial design using diol column.

Analyte	1	2	3	4	5	6	7	8	9	10	11
Acetone	2.23	2.19	2.23	2.19	2.23	2.19	2.23	2.19	2.18	2.20	2.20
Toluene	2.67	2.22	2.67	2.22	2.67	2.22	2.68	2.22	2.35	2.36	2.37

2'-Deoxyadenosine	2.22	3.52	2.21	3.51	2.22	3.57	2.21	3.51	2.49	2.55	2.55
2'-Deoxycytidine	2.06	3.53	2.01	3.57	2.05	3.73	2.01	3.66	2.29	2.33	2.34
2',3'-Dideoxyadenosine	2.49	3.77	2.47	3.73	2.48	3.78	2.47	3.69	2.74	2.82	2.82
2'-Deoxyguanosine	1.99	3.26	1.99	3.34	1.99	3.46	1.98	3.44	2.23	2.27	2.27
3'-Deoxyguanosine	1.99	3.15	1.99	3.22	1.99	3.34	1.98	3.32	2.22	2.25	2.25
5'-Methyluridine	1.94	2.35	1.94	2.37	1.94	2.41	1.94	2.41	2.06	2.08	2.08
8-Hydroxy-2'-deoxyguanosine	1.96	2.97	1.96	3.05	1.95	3.13	1.95	3.13	2.17	2.20	2.20
Adenosine	2.10	3.20	2.10	3.24	2.10	3.30	2.09	3.27	2.34	2.39	2.39
Cytidine	1.93	3.07	1.93	3.19	1.93	3.60	1.93	3.31	2.21	2.19	2.26
Guanosine	1.99	3.33	1.95	3.43	1.98	3.31	1.95	3.58	2.16	2.25	2.20
Inosine	1.93	2.78	1.93	2.88	1.92	2.95	1.92	2.98	2.13	2.16	2.16
Thymidine	1.99	2.37	1.99	2.38	1.99	2.41	1.99	2.41	2.10	2.12	2.12
Uridine	1.91	2.35	1.92	2.38	1.92	2.42	1.91	2.42	2.05	2.07	2.07
Adrenaline	2.27	4.89	2.83	6.50	2.21	4.86	2.51	6.16	2.92	3.10	3.15
Noradrenaline	2.23	4.86	2.71	6.35	2.17	4.86	2.41	5.91	2.83	2.98	3.02
3'-Methoxytyramine	2.42	5.46	3.05	6.77	2.37	5.35	2.71	6.42	3.12	3.33	3.37
Isoproterenol	2.34	4.40	2.97	5.53	2.28	4.22	2.64	4.97	2.90	3.04	3.07
Fenotrole	2.35	3.78	2.80	4.01	2.28	3.53	2.49	3.67	2.70	2.81	2.84
Terbutaline	2.35	4.06	2.85	4.71	2.29	3.83	2.53	4.29	2.80	2.93	2.96
Salbutamol	2.37	4.62	2.92	5.49	2.31	4.43	2.59	5.03	2.95	3.11	3.15
Ritudrine	2.48	4.11	3.02	4.51	2.41	3.81	2.68	4.06	2.84	2.98	3.01
Metaproterenol	2.32	4.23	2.81	4.99	2.26	4.03	2.50	4.56	2.80	2.94	2.97
Synephrine	2.36	5.17	2.97	6.58	2.31	5.04	2.64	6.35	3.04	3.24	3.29

Dopamine	2.31	5.05	2.88	6.42	2.26	4.99	2.55	6.05	2.95	3.12	3.16
N-methylephrine	2.71	5.75	3.85	7.31	2.69	5.53	3.41	6.74	3.65	3.96	4.03
Norphenylephrine	2.32	5.04	2.85	6.38	2.26	4.94	2.53	5.89	2.92	3.09	3.13
Phenylephrine	2.38	5.06	3.01	6.51	2.32	4.93	2.67	6.24	3.05	3.24	3.29
Tyramine	2.41	5.37	3.02	6.56	2.36	5.21	2.67	6.26	3.06	3.25	3.29
Normetanephrine	2.31	5.22	2.86	6.58	2.26	5.15	2.54	6.16	2.97	3.16	3.20
Octopamine	2.30	5.09	2.82	6.43	2.24	4.99	2.50	5.85	2.93	3.10	3.14
Methoxamine	2.60	5.39	3.34	6.47	2.55	5.17	2.94	6.10	3.24	3.46	3.50
Isoxuprine	2.73	4.25	3.30	3.70	2.67	3.92	2.97	3.47	2.93	3.12	3.14
Ceftiofur	1.96	2.51	1.68	2.60	1.98	2.65	-	2.74	1.91	1.91	1.90
Tyrosine	2.07	6.78	2.08	8.12	2.06	7.43	2.08	8.22	2.60	2.66	2.67
Pentoxyfylline	2.44	2.60	2.43	2.59	2.44	2.59	2.45	2.58	2.42	2.47	2.47
Guanine	2.13	3.84	2.12	3.96	2.13	4.04	2.12	4.05	2.43	2.49	2.48
Xanthine	2.02	2.67	2.01	2.78	2.02	2.76	2.02	2.84	2.19	2.22	2.22
Caffeine	2.40	2.65	2.39	2.64	2.40	2.65	2.40	2.63	2.44	2.48	2.48
Theophylline	2.25	2.66	2.25	2.68	2.25	2.67	2.26	2.69	2.35	2.38	2.38
Theobromine	2.22	2.65	2.22	2.65	2.22	2.67	2.22	2.66	2.33	2.37	2.37
Diphylline	2.10	2.67	2.10	2.69	2.10	2.71	2.10	2.70	2.25	2.29	2.29
7-Hydroxyethyltheophylline	2.21	2.63	2.20	2.63	2.21	2.64	2.21	2.63	2.31	2.35	2.35
1-Methyluric acid	1.99	2.62	1.77	3.82	1.99	2.69	1.84	3.99	2.17	2.18	2.17
1-Methylguanine	2.23	3.70	2.22	3.76	2.23	3.81	2.23	3.80	2.52	2.58	2.57
9-Methylguanine	2.22	3.64	2.21	3.71	2.21	3.80	2.21	3.78	2.48	2.55	2.55
Uric acid	1.91	2.75	1.67	4.49	1.91	2.89	1.76	4.81	2.09	2.10	2.09

3,7-Dimethyluric acid	2.03	2.65	1.85	3.90	2.03	2.71	1.93	4.05	2.28	2.30	2.29
7-Methylxanthine	2.10	2.67	2.10	2.70	2.10	2.72	2.11	2.73	2.25	2.28	2.29
Hypoxanthine	2.11	3.24	2.11	3.35	2.11	3.36	2.12	3.41	2.36	2.41	2.41
Proxyphylline	2.24	2.58	2.23	2.57	2.24	2.58	2.24	2.57	2.31	2.35	2.35
1,7-Dimethyluric acid	2.07	2.62	1.85	3.88	2.06	2.66	1.92	3.97	2.29	2.30	2.29
1,3-Dimethyluric acid	2.05	2.60	1.92	3.61	2.05	2.66	1.98	3.75	2.27	2.29	2.28
1,3,7-Trimethyluric acid	2.49	2.99	2.46	2.95	2.49	2.97	2.47	2.92	2.58	2.62	2.63
4-Nitrophenyl-B-D-glycopyranoside	2.00	2.27	2.00	2.30	2.00	2.32	2.00	2.33	2.05	2.08	2.08
Acyclovir	2.03	3.29	2.03	3.39	2.03	3.48	2.03	3.48	2.27	2.32	2.32
2'-Deoxyuridine	1.96	2.36	1.96	2.39	1.96	2.41	1.96	2.41	2.08	2.10	2.10
3'-Deoxythymidine	2.10	2.37	2.10	2.38	2.10	2.39	2.11	2.39	2.17	2.19	2.20
2'-Deoxyinosine	1.98	2.92	1.98	3.01	1.98	3.05	1.98	3.08	2.19	2.23	2.24
Satolol	2.40	4.32	2.99	5.01	2.34	4.10	2.65	4.59	2.91	3.07	3.11
7-Methylguanosine	2.24	6.43	2.59	8.49	2.19	6.66	2.37	8.03	3.02	3.22	3.30
Atenolol	2.41	5.79	3.04	7.08	2.36	5.61	2.70	6.57	3.16	3.38	3.45
Vadarabine	2.06	3.25	2.06	3.31	2.06	3.35	2.06	3.33	2.31	2.35	2.36
Tryptophan	2.28	7.07	2.28	8.21	2.27	7.56	2.28	8.26	2.81	2.90	2.92
BTMA	2.95	7.23	4.15	8.71	2.98	7.08	3.68	8.18	4.04	4.49	4.63
PTMA	2.90	7.44	4.09	9.02	2.93	7.32	3.63	8.48	4.07	4.54	4.70
Labetalol	2.74	4.43	3.54	4.88	2.68	4.11	3.10	4.40	3.13	3.32	3.37
Nadolol	2.47	5.48	3.15	6.74	2.42	5.30	2.78	6.39	3.18	3.38	3.45
Propranolol	2.96	5.20	4.09	6.37	2.92	4.87	3.58	5.68	3.57	3.83	3.90

Adenine	2.56	4.75	2.57	4.77	2.55	4.76	2.57	4.71	2.97	3.07	3.07
Uracil	2.03	2.39	2.03	2.40	2.03	2.41	2.04	2.42	2.13	2.16	2.17
Thymine	2.08	2.40	2.08	2.41	2.08	2.42	2.08	2.42	2.16	2.19	2.19
Cytosine	2.25	4.30	2.20	4.38	2.23	4.45	2.20	4.44	2.57	2.64	2.65
Pindolol	2.65	4.66	3.43	5.49	2.60	4.33	3.03	4.91	3.16	3.35	3.40
Alprenolol	2.89	4.83	3.91	5.69	2.85	4.50	3.43	5.14	3.41	3.64	3.70
Salicylic acid	2.08	2.28	1.74	2.22	2.09	2.34	-	2.30	1.92	1.92	1.91
5-Methylsalicylic acid	2.19	2.33	1.75	2.29	2.20	2.38	-	2.37	1.97	1.97	1.96
4-Hydroxybenzoic acid	2.15	2.22	1.93	3.03	2.15	2.23	1.99	3.12	2.20	2.23	2.24
3-Hydroxybenzoic acid	2.16	2.26	1.80	3.25	2.15	2.27	1.89	3.36	2.14	2.16	2.16
2,3-Dihydroxybenzoic acid	1.98	2.24	1.70	2.14	1.99	2.29	-	2.24	1.86	1.86	1.86
2,4-Dihydroxybenzoic acid	2.05	2.27	-	2.30	2.06	2.31	-	2.38	1.89	1.90	1.89
2,5-Dihydroxybenzoic acid	1.92	2.17	1.64	2.10	1.93	2.26	-	2.23	1.82	1.82	1.82
3,4-Dihydroxybenzoic acid	2.07	2.21	1.82	3.02	2.07	2.22	1.89	3.15	2.11	2.13	2.14
3,5-Dihydroxybenzoic acid	2.04	2.22	1.72	2.99	2.03	2.23	1.77	3.12	2.00	2.00	2.00
Benzoic acid	2.32	2.32	1.98	3.64	2.32	2.32	-	3.71	2.38	2.40	2.40
3-Amino-4-hydroxybenzoic acid	2.09	2.22	1.89	3.05	2.08	2.23	1.95	3.18	2.17	2.20	2.22
Gallic acid	2.00	2.20	1.72	3.19	1.99	2.22	1.81	-	2.04	2.05	2.05
4-Aminobenzoic acid	2.17	2.21	2.06	2.70	2.17	2.21	2.11	2.76	2.21	2.24	2.25
4-Aminosalicylic acid	2.16	2.26	1.73	2.47	2.15	2.28	-	2.55	1.96	1.96	1.96
Phthalic acid	1.82	1.88	1.66	2.94	1.85	2.00	1.76	3.59	1.84	1.86	1.88
Isophthalic acid	2.15	2.73	1.60	10.95	2.14	2.85	1.75	12.22	2.21	2.25	2.26
3-Aminobenzoic acid	2.19	2.28	1.88	3.58	2.19	2.28	1.97	3.70	2.30	2.31	2.32

Vanillic acid	2.17	2.25	1.93	3.25	2.16	2.25	2.01	3.34	2.27	2.29	2.29
Syringic acid	2.16	2.27	1.91	3.49	2.16	2.29	1.99	3.58	2.30	2.31	2.32
2-Methoxybenzoic acid	2.28	2.28	1.92	3.57	2.27	2.29	2.01	3.73	2.33	2.34	2.35
Terephthalic acid	2.15	2.78	-	9.71	2.10	2.86	1.74	10.65	2.23	2.26	2.27
Caffeic acid	2.11	2.23	1.90	3.31	2.10	2.24	1.98	3.44	2.23	2.24	2.24
P-Toluic acid	2.40	2.33	2.14	3.66	2.40	2.33	2.22	3.74	2.50	2.50	2.50
3,5-Dinitrosalicylic acid	-	1.58	1.67	1.53	-	1.63	-	1.58	1.69	1.67	1.68
Nicotinic acid	2.31	4.27	1.94	7.45	2.30	4.39	2.07	7.52	2.81	2.80	2.79
4-Hydroxybenzenesulfonic acid	1.66	1.87	1.56	1.82	1.70	1.99	1.65	1.92	1.70	1.69	1.69
4-Aminophenylacetic acid	2.16	2.27	1.98	4.07	2.15	2.29	2.07	4.28	2.45	2.46	2.49
Acetylsalicylic acid	2.22	2.31	1.80	3.26	2.22	2.33	1.91	3.33	2.18	2.18	-
Phenylacetic acid	2.26	2.26	-	3.96	2.26	-	2.20	4.07	2.53	2.53	2.54
1,2-benzenedisulfonic acid	-	2.70	-	2.78	-	3.27	-	3.18	-	1.76	1.76
P-Toluenesulfonic acid	1.75	1.87	1.61	1.82	1.75	-	1.71	1.91	1.74	1.73	-
2-Sulfobenzoic acid	1.73	2.25	-	5.22	1.73	2.68	-	6.09	2.50	1.93	1.95
Tropic acid	2.14	2.34	1.83	3.95	2.13	-	1.94	4.08	2.27	2.28	2.28
2-Phenylethylamine	2.64	5.86	3.49	7.04	2.59	5.68	3.08	6.57	3.44	3.62	3.69
Phenylalanine	2.25	7.36	2.26	8.63	2.24	7.89	2.27	8.70	2.85	2.93	2.96
Mandelic acid	2.10	2.68	1.74	3.04	2.08	-	1.86	3.15	2.08	2.08	2.08
5-Sulfosalicylic acid	-	2.08	1.49	2.07	-	-	1.55	-	1.64	1.62	1.62
Malic acid	1.74	1.82	1.60	1.88	1.75	1.92	1.71	2.04	1.74	1.73	1.73

Numbers in row are operating conditions of the full factorial design (Table 2.3).

R2013b (The Mathworks Inc., Natick, MA, USA) was utilised to select the most important descriptors and to build the QSRR models for each stationary phase material. Statistical evaluation of the data and multivariate data analysis has also been performed in Matlab.

2.2.2 Calculation of molecular descriptors

The procedure for the generation of molecular descriptors in this study was as follows. The structures of the molecules were sketched in MarvinSketch. Initial conformational searches to find the 50 lowest energy structures were performed using Balloon with a Merck Molecular Force Field (MMff94) [6-9]. The lowest energy conformer was taken as the input structure for geometry optimisation using a semi-empirical PM7 method implemented in MOPAC [12], the resulting geometry was further refined with the Gaussian program applying the Becke 3-parameter (exchange) with correlation by Lee Yang and Parr, (B3LYP) [18-21] functional and the 6-31G-(d) basis set [22]. Optimisations were performed in acetonitrile using the integral equation formalism variant of the polarisable continuum model (IEFPCM) [23]. Following each geometry optimisation, harmonic frequency analysis was carried out to confirm the nature of each stationary point as an equilibrium structure.

The resulting minimum energy conformations of the compounds in this study were input into Dragon to calculate molecular descriptors. Dragon software [16] was able to calculate 2,687 molecular descriptors, consisting of constitutional, topological, geometrical, electrostatic, physical, shape, and quantum chemical descriptors. The Handbook of Molecular Descriptors [24] details the calculation procedure.

To minimize subsequent problems of chance correlation, descriptors with constant or near constant values, descriptors with a standard deviation less than 0.0001, descriptors which were strongly correlated to other descriptors (using a correlation coefficient >0.90) and those descriptors not available for all compounds were excluded. After this reduction step, 321 molecular descriptors were obtained. Before statistical analysis, all the descriptors were scaled to zero mean and unit variance (auto-scaling procedure) because the numerical values of the descriptors varied significantly. The resulting descriptor sets were used to build predictive models for the experimental chromatographic retention data.

2.2.3 Genetic algorithm (GA)

The GA, introduced by Holland [25], is a stochastic search procedure inspired by the rules of natural selection to select features without making any assumptions about the search space. The foundation of the procedure is based on assignment of greater reproductive opportunities to solutions that have higher fitness. In genetic terms, each variable is called a gene (bit), and a set of variables is called a chromosome (bit string). In the first step of the GA optimisation, an initial population of chromosomes is generated by a random choice of each variable. Then, pairs of chromosomes are chosen randomly from the original population as parents and crossover operations performed to produce a new generation of child chromosomes with better fitness. The last step involves a mutation process to maintain genetic diversity from the initial random population to the next generations. The cycle of the evaluation, selection, crossover, and mutation processes is then repeated until a stopping criterion is satisfied.

One limitation of the GA is that it relies on a randomly generated initial population, which can potentially limit its capability to find the most relevant variables within a large search domain. As a consequence the final

results of replicate runs can be substantially different. To reach a suitable subset of descriptors within a reasonable computational time, 100 runs with different initial populations were generated and the frequency with which each variable was selected as the top chromosome of each run was calculated and PLS regression was used to determine the set of the most relevant descriptors, and to create the final model. More details of this approach are available elsewhere [26, 27]. This approach is termed GA-PLS.

2.2.4 Partial least square regression (PLS)

PLS regression was employed as a multivariate method to decrease the dimensionality of the large set of independent molecular variables by extracting a small number of latent variables (see below) that are correlated with the dependent variable (i.e., chromatographic retention time). In addition, PLS rotates the latent variables by maximising the covariance between molecular descriptors and the dependent variable. This ensures that the molecular descriptors which are highly correlated to the dependent variable are retained in the first few latent variables. The PLS method is presented in equations (1.9) and (1.10).

The optimum number of latent variables for each model was selected using 4-fold cross validation. PLS models with a number of latent variables up to 5 were investigated and the optimum number of LVs in each model was selected by applying the first standard deviation rule [28, 29] to avoid overfitting.

2.2.5 Model validation

The first step in the modelling involves defining a training set for calibration and a test set for validation. It has also been emphasized that an external set is required to validate the predictive power of a quantitative-

property model [30-32]. For this purpose, we applied an independent set, separate from the training samples. The training set was used to select the descriptors for QSRR modelling and to build the models. Then the experimental chromatographic data of the test compounds were collected and compared with their predicted data calculated from derived QSRR models.

The final models were evaluated for their predictive ability using the mean absolute error (MAE) and the root-mean-square error prediction (RMSEP) defined [33] as

$$MAE = \frac{|y_i(obsd) - y_i(pred)|}{n} \qquad (2.1)$$

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n}\left(\frac{y_i(obsd) - y_i(pred)}{yi(obsd)}\right)^2}{n}} \times 100\ (\%) \qquad (2.2)$$

where yi(obsd) and yi(pred) are the observed and predicted retention times and n is the number of analytes. The calibration models were examined for their quality by root-mean-square error cross validation (RMSECV) in percentage and absolute terms, and the squared correlation coefficient Q^2 defined as

$$RMSECV(\%) = \sqrt{\frac{\sum_{i=1}^{n}\left(\frac{y_i(obsd) - y_i(pred)}{yi(obsd)}\right)^2}{n}} \times 100\ (\%) \qquad (2.3)$$

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{n}(yi(obsd) - yi(pred))^2}{n}} \qquad (2.4)$$

$$Q^2 = 1 - \frac{\sum_{i=1}^{n}\left(yi(obsd) - yi(pred)\right)^2}{\sum_{i=1}^{n}\left(yi(obsd) - y(mean)\right)^2} \qquad (2.5)$$

determined from the predicted retention times yi(pred) of the test set analyte(s) (*i.e.* those analytes left out of the training set) during cross validation. The y(mean) is the average value of the observed retention times.

Model validation includes Y-randomisation tests to evaluate the statistical significance of the estimated predictive power based on a response variable randomisation process. For this purpose, the response variable was randomised, and a 4-fold cross-validation was performed with the entire model development procedure [34]. As our data sorted in the real order gave better prediction results that randomly permuted data, it can be concluded that the prediction model is significant.

2.3 References

[1] G. Schuster, W. Lindner, Comparative characterisation of hydrophilic interaction liquid chromatography columns by linear solvation energy relationships, J. Chromatogr. A, 1273 (2013) 73-94.

[2] A. Periat, B. Debrus, S. Rudaz, D. Guillarme, Screening of the most relevant parameters for method development in ultra-high performance hydrophilic interaction chromatography, J. Chromatogr. A, 1282 (2013) 72-83.

[3] A. Kumar, J.C. Heaton, D.V. McCalley, Practical investigation of the factors that affect the selectivity in hydrophilic interaction chromatography, J. Chromatogr. A, 1276 (2013) 33-46.

[4] Y. Guo, S. Gaiki, Retention and selectivity of stationary phases for hydrophilic interaction chromatography, J. Chromatogr. A, 1218 (2011) 5920-5938.

[5] MarvinSketch, in, ChemAxon, http://www.chemaxon.com [accessed January 2016].

[6] T.A. Halgren, R.B. Nachbar, Merck molecular force field. IV. conformational energies and geometries for MMFF94, J. Comput. Chem., 17 (1996) 587-615.
[7] T.A. Halgren, Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94, J. Comput. Chem., 17 (1996) 553-586.
[8] T.A. Halgren, Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions, J. Comput. Chem., 17 (1996) 520-552.
[9] T.A. Halgren, Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94, J. Comput. Chem., 17 (1996) 490-519.
[10] M.J. Vainio, M.S. Johnson, Generating conformer ensembles using a multiobjective genetic algorithm, J. Chem. Inf. Model., 47 (2007) 2462-2474.
[11] J.J. Stewart, Optimisation of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimisation of parameters, J. Mol. Model., 19 (2013) 1-32.
[12] MOPAC2012, in, Stewart, J. J. P. Stewart Computational Chemistry, Colorado Springs, CO, USA, HTTP://OpenMOPAC.net [accessed October 2015].
[13] R. Parr, W. Yang, Density-Functional Theory of Atoms and Molecules, Oxford University Press, New York, 1989.
[14] W. Koch, M.C. Holthausen, A Chemist's Guide to Density Functional Theory, Wiley-VCH, Weinheim, Germany, 2001.
[15] M.J.T. Frisch, G. W. Schlegel, H. B. Scuseria, G. E. Robb, M. A. Cheeseman, J. R. Scalmani, G. Barone, V. Mennucci, B. Petersson, G. A. Nakatsuji, H. Caricato, M. Li, X. Hratchian, H. P. Izmaylov, A. F. Bloino, J. Zheng, G. Sonnenberg, J. L. Hada, M. Ehara, M. Toyota, K. Fukuda, R.

Hasegawa, J. Ishida, M. Nakajima, T. Honda, Y. Kitao, O. Nakai, H. Vreven, T. Montgomery, Jr., J. A. Peralta, J. E. Ogliaro, F. Bearpark, M. Heyd, J. J. Brothers, E. Kudin, K. N. Staroverov, V. N. Kobayashi, R. Normand, J. Raghavachari, K. Rendell, A. Burant, J. C. Iyengar, S. S. Tomasi, J. Cossi, M. Rega, N. Millam, J. M. Klene, M. Knox, J. E. Cross, J. B. Bakken, V. Adamo, C. Jaramillo, J. Gomperts, R. Stratmann, R. E. Yazyev, O. Austin, A. J. Cammi, R. Pomelli, C. Ochterski, J. W. Martin, R. L. Morokuma, K. Zakrzewski, V. G. Voth, G. A. Salvador, P. Dannenberg, J. J. Dapprich, S. Daniels, A. D. Farkas, O. Foresman, J. B. Ortiz, J. V. Cioslowski, J. and Fox, D. J. Gaussian 09, Revision A.02,, in, Gaussian, Inc., Wallingford CT, 2009.

[16] Dragon 6.0 for Windows (Software For Molecular Descriptor Calculations); http://www.talete.mi.it/ Talete, Milano, Italy [accessed December 2015].

[17] Matlab, in The Mathworks Inc., Natick, MA, USA, 2013.

[18] A.D. Becke, A new mixing of Hartree–Fock and local density-functional theories, J. Chem. Phys., 98 (1993) 1372.

[19] A.D. Becke, Density-functional exchange-energy approximation with correct asymptotic behaviour, Phys. Rev. A, 38 (1988) 3098-3100.

[20] C. Lee, W. Yang, R.G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density, Phys. Rev. B, 37 (1988) 785-789.

[21] P.J. Stephens, F.J. Devlin, C.F. Chabalowski, M.J. Frisch, Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields, J. Phys. Chem., 98 (1994) 11623-11627.

[22] M.J. Frisch, J.A. Pople, J.S. Binkley, Self-consistent molecular orbital methods. 25. Supplementary functions for Gaussian basis sets, J. Chem. Phys., 80 (1984) 3265-3269.

[23] J. Tomasi, B. Mennucci, R. Cammi, Quantum mechanical continuum solvation models, Chem. Rev., 105 (2005) 2999-3093.
[24] R. Todeschini, V. Consonni, Handbook of Molecular Descriptors, Wiley-WCH, Weinheim, 2000.
[25] J.H. Holland, Adaptation in Natural and artificial Systems, MI, University of Michigan Press: Ann Arbor, 1975.
[26] R. Leardi, Application of genetic algorithm-PLS for feature selection in spectral data sets, J. Chemom., 14 (2000) 643-655.
[27] R. Leardi, A. Lupiáñez González, Genetic algorithms applied to feature selection in PLS regression: how and when to use them, Chemom. Intell. Lab. Syst., 41 (1998) 195-207.
[28] K. Varmuza, P. Filzmoser, M. Dehmer, Multivariate linear QSPR/QSAR models: Rigorous evaluation of variable selection for PLS, Comput. Struct. Biotechnol. J., 5 (2013) e201302007.
[29] K. Varmuza, P. Filzmoser, Introduction to multivariate statistical analysis in chemometrics, USA: CRC Press, Boca Raton, FL, 2009.
[30] A. Tropsha, P. Gramatica, V.K. Gombar, The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models, QSAR Comb. Sci., 22 (2003) 69-77.
[31] P. Gramatica, Principles of QSAR models validation: internal and external, QSAR Comb. Sci., 26 (2007) 694-701.
[32] V. Consonni, D. Ballabio, R. Todeschini, Evaluation of model predictive ability by external validation techniques, J. Chemom., 24 (2010) 194-201.
[33] P. Zuvela, J.J. Liu, K. Macur, T. Baczek, Molecular descriptor subset selection in theoretical peptide quantitative structure-retention relationship model development using nature-inspired optimisation algorithms, Anal. Chem., 87 (2015) 9876-9883.

[34] C. Rucker, G. Rucker, M. Meringer, y-Randomization and its variants in QSPR/QSAR, J. Chem. Inf. Model., 47 (2007) 2345-2357.

3 Prediction of retention in hydrophilic interaction liquid chromatography using solute molecular descriptors based on chemical structures

3.1 Introduction

HILIC [1] has become an important alternative to reversed-phase liquid chromatography (RPLC) for polar analytes. Recent developments have meant that this method is now well-recognized as a powerful and selective technique, which has been employed successfully for the separation of numerous polar compounds [2, 3] including β-adrenergic agonists and related compounds [4-8]. The availability of a broad range of HILIC stationary phases provides opportunities for meaningfully different retention and separation selectivity. As a consequence, many mixtures of polar compounds may be separated by means of HILIC. However, an outcome of the development of new stationary phases is that it is now more challenging to select the most suitable stationary phase.

In order to screen stationary phases, trial and error optimisation is used frequently, although this may take many attempts, with subsequent loss of time. Another approach, based on column characterisation and classification methods, guides users to identify either similar or diverse stationary phases depending on method development needs [5, 9, 10]. However, the application of these methods is limited only to those analytes having available experimental descriptors or molecules with properties somewhat comparable to those studied. In addition, because of limited understanding of the HILIC mechanism, the quality of these approaches, which depend strongly on *a priori* knowledge of the retention mechanism, is doubtful. Consequently there is a strong demand for a powerful tool to handle stationary phase selection based on more objective criteria. Statistically-

derived quantitative structure-retention relationships (QSRRs) represent a popular chemometric approach in High-Performance Liquid Chromatography (HPLC) method development [11] and may be put into practice to accelerate the screening of stationary phases in liquid chromatography method development by predicting retention of target analytes across available chromatographic columns.

QSRR is a powerful theoretical tool capable of prediction of the chromatographic behaviour of a given chromatographic system, which then can be used for future retention predictions of new compounds. The aim of developing a QSRR model is to construct a statistically significant mathematical relationship between a chromatographic parameter (eg., retention time) and some molecular descriptors characterising the molecular structure of the analytes. Molecular descriptors are either determined experimentally or computed theoretically. QSRR studies start from the calculation and selection of appropriate descriptors, followed by regression analysis to derive mathematical models of retention parameters as a function of these descriptors.

A wide variety of analyte descriptors has been used in QSRR studies, ranging from physicochemical properties of analytes (such as molecular weight, polar surface area, logP, logD, etc) to molecular descriptors calculated from chemical structures optimised using density functional theory (DFT). Advances in computational chemistry permit the calculation of more than 4000 theoretical descriptors for an individual analyte, based only on chemical structure, which can be used in QSRR modelling [12]. Some of these descriptors may be redundant, be irrelevant, or represent noise. Thus, for proper QSRR modelling an appropriate method of selection of the most important descriptors prior to regression analysis is crucial to eliminate unnecessary descriptors. Although several such variable selection

techniques and regression methods have been applied in QSRR studies in various modes of chromatography [13, 14], the current work is based on the use of a partial least squares modelling approach (PLS) with a genetic algorithm (GA) employed as a widely and successfully utilised approach to variable selection [15, 16]. PLS has been chosen as a regression method for QSRR modelling because of its utility in handling large sets of collinear descriptors, as well as having small demands on computational time and effort [17].

Only a few reports demonstrating the application of QSRR techniques for prediction of retention of analytes in HILIC have been published. Kaliszan et al. [18] and Burgess et al. [19] constructed linear models for the prediction of metabolite retention times in HILIC for the purpose of removing false identification during the interpretation of metabolomics data. Jinno et al. [20] reviewed the development of retention prediction models of adrenoreceptor agonists and antagonists in HILIC systems using selected molecular descriptors. More recently, Cao et al. [21] established a QSRR model based on the Random Forest algorithm for the prediction of retention time in HILIC, allowing peak annotation of metabolites. However, a study of the application of a GA as a variable selection method in QSRR methodology for HILIC method development has, to the best of our knowledge, not yet been published. Therefore, the main aim of this study, which is part of a broader structure-retention relationship design project, was to establish a QSRR model based on molecular descriptors computed from chemical structures optimised using DFT for the prediction of retention of a class of compounds for five different HILIC stationary phases utilising a GA coupled with PLS for variable selection. The second aim of this work was to present for the first time a strategy to optimise the GA descriptor sets in QSRR models and to compare the performance of this optimised approach with more common approaches. Finally, in order to

obtain some insight into the HILIC mechanism, the selected molecular descriptors in the QSRR models have been investigated in each HILIC system.

This study has been undertaken using β-adrenergic agonists as one classification of analytes. β-adrenergic agonists are structurally related compounds with low log D values presenting hydrophilic properties. These species are synthetic phenethanolamine compounds used as bronchodilator and tocolytic agents in human as well as in veterinary medicine. Furthermore, they are employed by the livestock industries to improve feed efficiency and growth rates [22, 23]. However, they may provide human health risks [22] because of their growth-promoting effects. As a result, β-adrenergic agonists should be monitored and controlled strictly. European Union countries [24] have prohibited β-adrenergic agonists as growth promoters in food producing animals according to toxicological data. Reversed phase liquid chromatography (RPLC) coupled to mass spectrometry has been the most popular analytical method to monitor the illicit usage of β-adrenergic agonists and related compounds [25-28]. However, these methods face some issues related to co-elution of analytes resulting from the high polarity of these compounds. HILIC is therefore the preferred separation mode, but selection of the suitable stationary phase, as a starting point for method development for simultaneous determination of β-adrenergic agonists, constitutes a significant analytical challenge.

3.2 Method

Five data sets composed of a training set made up of 16 analytes and a test set with 6 analytes with experimental retention times over five HILIC stationary phases (bare silica, amine, amide, diol and zwitterionic) were used. These data involved β-adrenergic agonists and related compounds, as seen in Table 3.1 and Figure 3.1. The isocratic eluent contained 90:10

acetonitrile−formate buffer solution. Formate buffer 100 mM was prepared with an adapted volume of ammonium formate and the pH adjusted to 3.0 with formic acid. Details of collected data are presented in Chapter 2.

The retention times observed on each stationary phase are provided in Table 3.1.

For the modelling step, an unmodified GA-PLS method was used and the details of this approach are stated in Chapter 2. As a slightly modified approach, the GA-PLS procedure was performed 10 times and those descriptors which appeared in every iteration (i.e., with 100% selection frequency) were chosen as an optimum subset of descriptors. This set of descriptors was then used to build the final PLS model for retention prediction. This approach is termed Optimised GA-PLS. For reference purposes, PLS was also applied to the full set of descriptors without the use of any variable selection method. This approach is termed Full PLS. These three modelling approaches are shown schematically in Figure 3.2.

3.3 Result and discussion

3.3.1 Analytes and retention behaviour

Figure 3.1 shows the chemical structures of the β-adrenergic agonists studied. Chromatographic retention prediction of these compounds was evaluated on five HILIC stationary phases: amide, amine, diol, bare silica and zwitterionic (see Chapter 2 for definitions). These phases were chosen primarily based on differences in their chemistries to reasonably cover the HILIC chromatographic space (Figure 3.3). The amine and zwitterionic columns carry charged functionalities, whereas the diol, amide and bare silica columns are neutral under the chromatographic conditions used in this study.

Table 3.1. The experimental retention times of the training and test set on five different HILIC stationary phases.

Column / Analyte	Diol	Bare silica	Amine	Amide	Zwitterionic
Training set	$tR_{exp.}$	$tR_{exp.}$	$tR_{exp.}$	$tR_{exp.}$	$tR_{exp.}$
adrenaline	4.89	6.73	10.82	7.75	17.30
noradrenaline	4.86	7.30	14.71	9.81	24.97
3-methoxytyramine	5.46	4.86	6.17	5.98	9.30
isoproterenol	4.40	4.63	7.13	6.40	10.57
fenotrole	3.78	2.72	5.11	6.75	9.44
salbutamol	4.62	4.59	6.51	6.49	9.47
ritudrine	4.11	2.47	3.77	5.43	5.72
metaproterenol	4.23	4.10	6.63	6.41	11.52
synephrine	5.17	5.20	6.84	6.32	10.59
dopamine	5.05	5.62	9.23	7.64	15.02
norfenefrine	5.04	5.26	8.81	7.58	15.24
phenylephrine	5.06	4.92	6.73	6.06	10.41
normetanephrine	5.22	5.85	8.79	7.55	14.14
isoxuprine	4.25	2.11	2.87	3.98	3.65
etilefrine	6.43	4.11	4.99	4.67	8.34
N-methylephedrine	5.75	3.90	3.80	3.98	4.73
Test set					
terbutaline	4.06	3.53	5.71	5.88	9.57
tyramine	5.37	4.54	6.18	6.26	9.55
octopamine	5.09	5.52	8.93	7.93	14.85
metanephrine	6.29	5.54	5.89	5.73	10.67
phenylethylamine	5.43	3.98	4.02	5.27	6.51
3-methylphen ethylamine	6.50	3.61	3.59	4.40	5.72

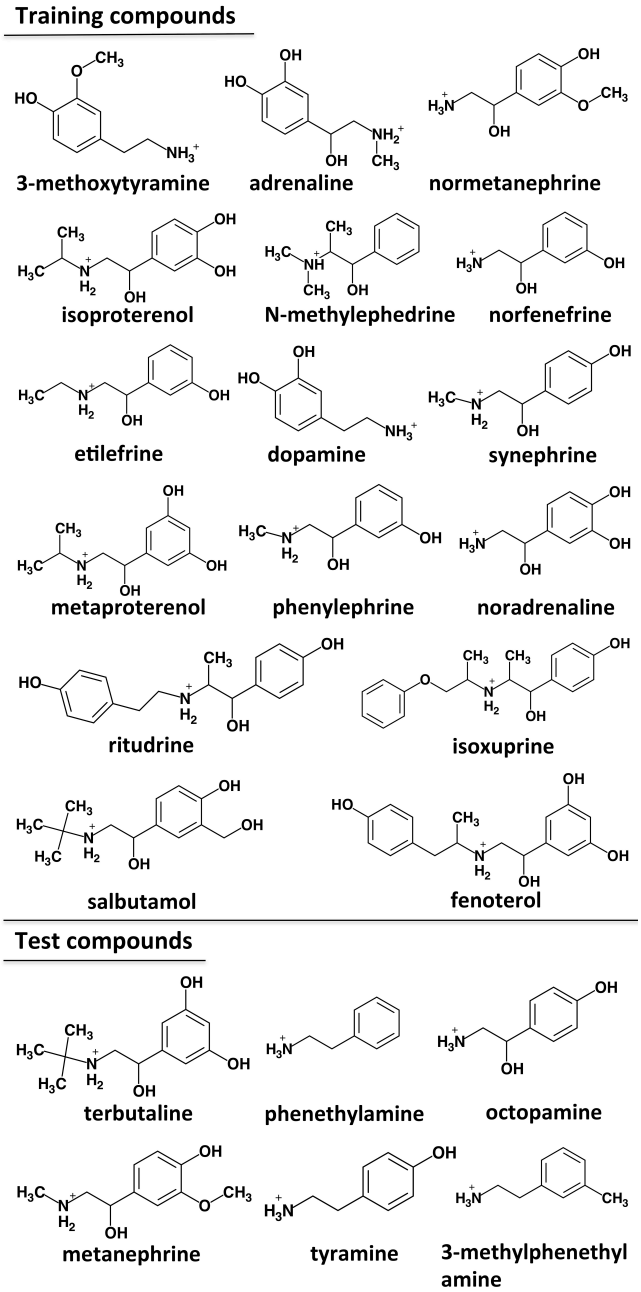

Figure 3.1. Structures of the studied β-adrenergic agonists and related compounds.

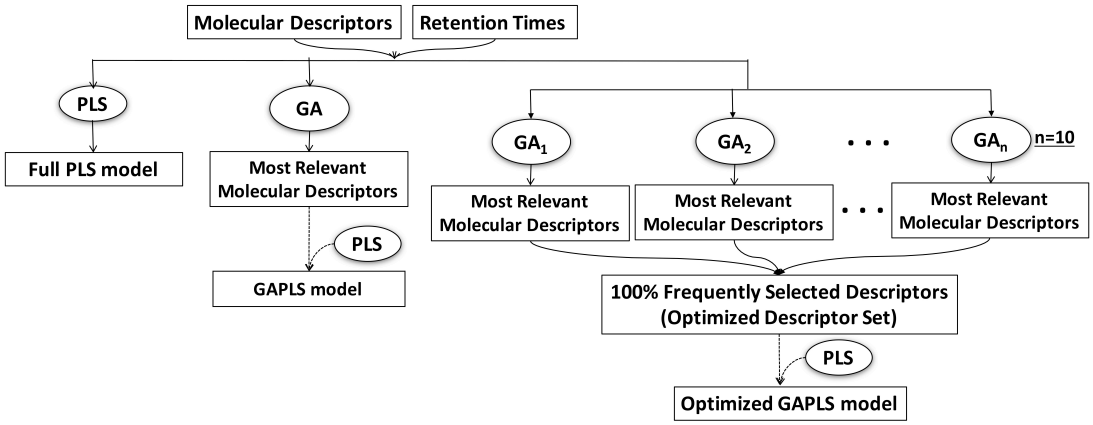

Figure 3.2. Scheme of QSRR modelling in this study.

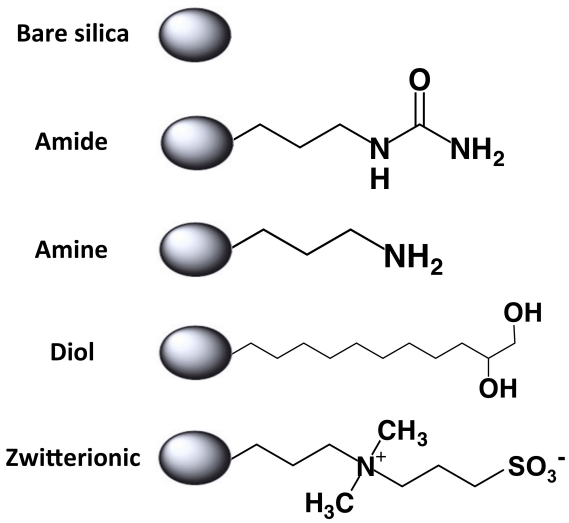

Figure 3.3. Structures of the Thermo Fisher HILIC stationary phases studied.

A comparison of the retention factors of the β-adrenergic agonists on the five stationary phases is shown in Figure 3.4. As seen in the figure, the retention factors of the β-adrenergic agonists are increased on the amide column compared to the diol column, indicating higher hydrophilic partitioning. In addition, there are cases where the order of elution differed between these two neutral stationary phases. One such pair of molecules is etilefrine and ritudrine. An increased retention factor in the bare silica system was observed compared to neutral columns, which may be partially attributable to the electrostatic attraction between the positively charged basic and the possibly negatively charged silanol groups [5]. Another observation is a significant increase in the retention factor of almost all β-adrenergic agonists when the stationary phase type was changed from a neutral column to a zwitterionic column, due possibly to electrostatic attraction between the positively charged bases and the negative charges on the surface of zwitterionic stationary phase. Another interesting result was observed for the amine phase, where a general increase in the retention factor of β-adrenergic agonists compared to neutral phases was noted. These results suggested that the anticipated electrostatic repulsion between the positively charged β-adrenergic agonists and the positively charged amino phase was significantly diminished, possibly due to interaction of the formate counterion in the eluent with the amine group of the surface. Our result is consistent with the literature in which it can be seen that a high salt concentration (10 mM formate) has been shown to promote stronger hydrophilic partitioning [57]. This would account for the observation that basic β-adrenergic agonists are better retained in the positively charged amine column compared with the neutral columns. These experimental results indicated that significant changes in retention and separation selectivity can occur depending on the stationary phase chemistry.

3.3.2 QSRR modelling

QSRR models based on the three variable selection approaches (see Figure 3.2) were generated for each HILIC system. These models were generated using the experimental chromatographic retention times for sixteen β-adrenergic agonists - the training set - on five HILIC stationary phases, (see Table 3.1) together with molecular descriptors calculated for these analytes. Also included in Table 3.1 are the observed retention times for the six compounds comprising the analyte test set. The aim of this work was to generate models that could predict β-adrenergic agonist retention times directly from molecular structure data and that could also assist in explaining the main interactions that take place in HILIC systems. External validation of the predictive power of the different QSRR models was evaluated using the test analytes, which were not utilised in either descriptor feature selection or model generation. The appropriate descriptor values selected by the modelling approach were inserted into the correlation equation, and the respective retention times were calculated and compared to measured experimental retention times. The predicted retention times of the test set of analytes are shown in Table 3.2, for each of the three modelling approaches. A summary of the overall performance of each modelling approach for each stationary phase is given in Table 3.3. The number of descriptors included in the models ranged from 321 (for the Full PLS model) to 6 (for the Optimised GA-PLS model on the diol column). The number of descriptors used for modelling decreased in the order Full PLS > GA-PLS > Optimised GA-PLS. Although the relevance of some of the chosen descriptors is still not completely clear in terms of their significance to the HILIC retention process, other descriptors are clearly linked with chemical properties that are relevant to the HILIC experimental system. These implied relationships are discussed in more detail in the following section.

Table 3.3 lists the mean absolute error (MAE, s) for each modelling approach and stationary phase, and this error was consistently lowest for the optimised- GA-PLS approach, despite the fact that this approach used the fewest descriptors. This shows clearly that it is the relevance of the descriptors used for modelling, not the number of descriptors, which determines the predictive quality of the final model.

Table 3.3 also lists the errors as RMSEP (see Eqn. 2.4) which scales the errors to the observed retention time of each analyte and is expressed as a percentage. The RMSEP for the optimised GA-PLS approach was again consistently the lowest of the three modelling approaches. The obtained results indicate that the optimised GA-PLS approach provided an acceptable level of accuracy of retention time prediction using a relatively small descriptor set, while not requiring excessive computational resources (for example, compared to alternative approaches for example reference [16]). The average CPU time for optimised GA-PLS modelling was 381 s. Figures 3.5 and 3.6 show correlation plots between observed and predicted retention times.

To the authors' knowledge, this is the first published application of GA-PLS with an optimised descriptor set in QSRR, and GA feature selection in general in the HILIC mode. While some retention time prediction studies have been made through QSRR models in HILIC [20, 29-31], it is difficult to compare the accuracy of these models since the same prediction error statistics are not always reported. One QSRR model has been reported with relative errors of prediction more than 100% for an eight-min range of retention times [18]. Better accuracy has been achieved for adrenoreceptor agonists and antagonists on three HILIC systems using artificial neural networks and multiple linear regression to develop QSRR models based on

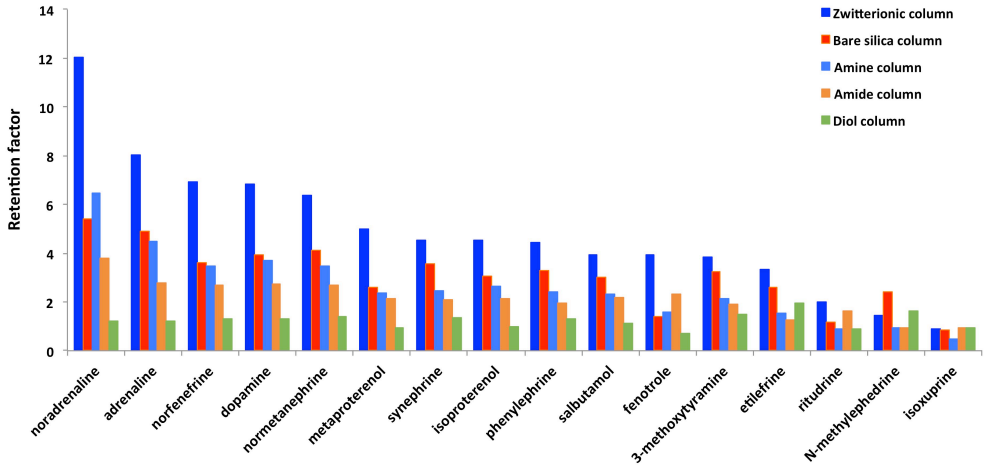

Figure 3.4. Retention factors of β-adrenergic agonists on zwitterionic, amine, amide, diol and bare silica HILIC stationary phases.

Table 3.2. The predicted retention times of the test compounds on five different HILIC stationary phases applying optimised GA-PLS, GA-PLS and full PLS.

Column / Test analyte	Diol	Bare silica	Amine	Amide	Zwitterionic
Optimised GAPLS	$tR_{Pred.}$	$tR_{Pred.}$	$tR_{Pred.}$	$tR_{Pred.}$	$tR_{Pred.}$
terbutaline	4.11	3.31	6.08	6.10	10.37
tyramine	5.14	4.30	5.76	5.91	9.42
octopamine	5.02	6.05	9.51	7.83	15.66
metanephrine	6.14	5.24	5.79	5.78	9.25
phenylethylamine	6.04	3.75	3.59	4.92	5.29
3-methylphen ethylamine	6.28	3.62	4.21	5.16	6.32
GAPLS					
terbutaline	4.20	3.77	5.75	6.36	9.39
tyramine	5.16	4.40	7.50	6.00	10.37
octopamine	5.14	5.77	10.40	7.85	16.32
metanephrine	5.18	4.88	6.00	5.91	9.34
phenylethylamine	5.42	3.66	3.21	4.94	8.45
3-methylphen ethylamine	5.63	3.72	3.35	5.44	6.90
Full PLS					
terbutaline	4.39	3.90	5.71	6.51	9.02
tyramine	5.23	5.31	8.11	6.31	13.00
octopamine	5.23	5.94	9.79	7.73	16.05
metanephrine	5.06	4.81	6.88	5.90	10.68
phenylethylamine	5.53	5.20	7.33	4.95	11.48
3-methylphen ethylamine	5.42	4.98	6.90	5.14	10.68

Table 3.3. QSRR models performance summary

Approach / Column (a, b)	nVar	MAE (s)	RMSEP
Diol (4.90, 0.67)			
Optimised GA-PLS	6	13	4.88
GA-PLS	43	24	9.28
Full PLS	321	30	11.13
Bare silica (4.65, 1.43)			
Optimised GA-PLS	17	15	6.07
GA-PLS	63	17	6.94
Full PLS	321	49	22.25
Amine (7.06, 2.97)			
Optimised GA-PLS	10	25	9.55
GA-PLS	22	40	14.03
Full PLS	321	78	52.56
Amide (6.42, 1.50)			
Optimised GA-PLS	16	19	8.06
GA-PLS	17	24	10.75
Full PLS	321	21	8.67
Zwitterionic (11.28, 5.26)			
Optimised GA-PLS	9	50	11.12
GA-PLS	25	69	16.56
Full PLS	321	151	49.60

Method abbreviations explained in the text. nVar is the number of selected variables (descriptors). a and b are the average retention and standard deviation, respectively, observed in the training set.

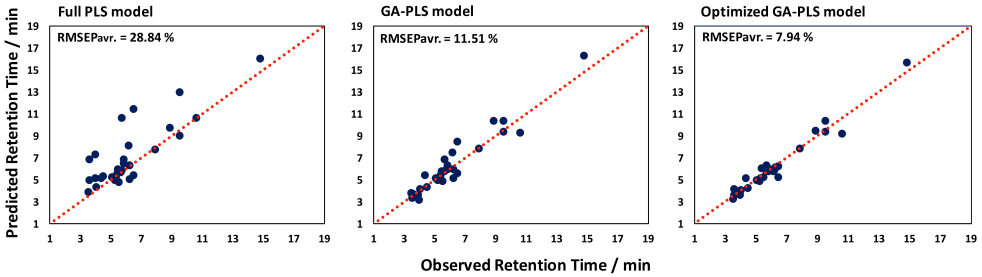

Figure 3.5. Predictive ability of optimised GA-PLS models, best GA-PLS models and full PLS models for an external validation set of β-adrenergic agonists over five different HILIC systems. RMSEPavr. is the average value of RMSEP of test β-adrenergic agonists over five HILIC stationary phases.

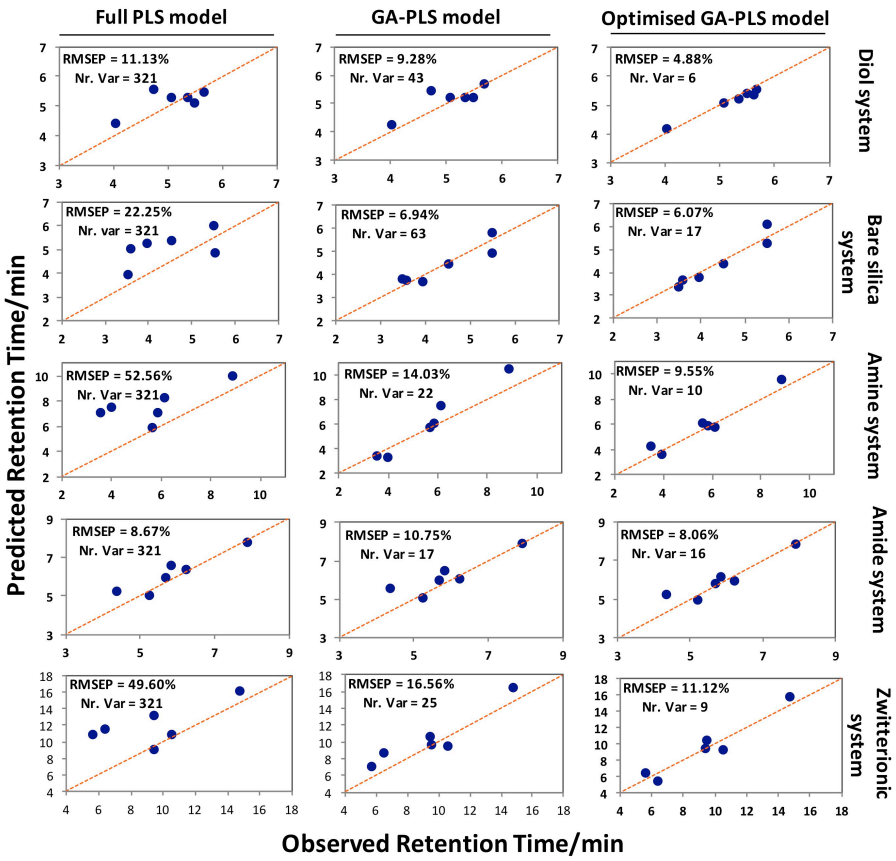

Figure 3.6. Predictive ability of optimised GA-PLS models, best GA-PLS models and full PLS models for an external validation set of β-adrenergic agonists over five different HILIC systems.

pre-defined descriptors [29-31]. All of the analysed compounds in these studies had retention factors less than one for all reported chromatographic conditions [29-31]; therefore the models were constructed with compounds that interacted only weakly with the chromatographic system. In contrast, the retention factors in our study range from 0.46 to 12.02 with the majority of retention factors being greater than 1 (see Figure 3.4). Our results from the QSRR modelling section indicate that the optimised GA-PLS models generated for these experimental systems were very well correlated with the experimental data, reliable, and strongly robust, and are capable of making good *a priori* predictions for β-adrenergic agonists. To illustrate this, Figure 3.7 shows the predicted retention times of the six test analytes (indicated by circles) on five HILIC columns. The level of accuracy of prediction is sufficient for the QSRR approach to be used to determine which of the five HILIC stationary phases are likely to yield a suitable best separation. Further experimental studies would then be necessary to identify the precise composition of the optimal eluent.

To test the reliability of the built QSRR models, a number of plots were generated. First, variable importance to projection (VIP) [32] plots were used (Figure 3.8) to gain an understanding of the relative importance of each descriptor for the GA-PLS models being generated, based on optimised molecular descriptors. As seen in the figure, some molecular descriptors had VIP values lower than the threshold of 1 [33-35] but were retained in order to balance the multivariate models.

Second, a complete correlation map between the descriptors and the response (i.e. retention time) (Figure 3.9) showed the presence of collinear descriptors. The presence of a number of collinear descriptors in the final model is very common and may play an important role in adding reliability to QSRR models [36].

Next, to demonstrate the absence of chance correlation in the QSRR models, Y-randomization tests (Figure 3.10) each with 1000 iterations were applied with randomly assigned retention times to the training sets. For each iteration, a GA-PLS calculation with an optimised descriptor set was performed on the permuted retention time data. This yielded average cross-validated RMSE values of 25 – 69% for the five columns. This further confirmed the predictive ability of the models, given that RMSE values for the actual dataset were shown to be significantly lower than those obtained for the same dataset with randomized retention data.

Finally, an applicability domain of the optimised GA-PLS models and the reliability of predictions was evaluated by the leverage approach expressed as a Williams plot [37]. In the Williams plot, standardised residuals versus leverage values hi are plotted. The leverage value (hat value), hi, was defined as: $hi = xi^T(X^TX)xi$, where xi is the descriptor vector of the considered compound and X is the descriptor matrix derived from the training set descriptor values. A leverage greater than the critical leverage value h^* warns the potential extrapolation of the model and the predicted response may not be reliable. A critical leverage value was determined as: $h^* = 3p/n$, in which n is the number of observations used to generate the model and p is the number of parameters in the model. The standardised residual reflects the prediction quality, and the acceptable range is usually (±3σ). Figure 3.11 presents the Williams plots of the compounds under study with ±3σ, and h^* (0.5, 1.6, 1.0, 1.5 and 0.9 for diol, bare silica, amine, amide and zwitterionic systems, respectively) as warning limits. It is obvious that only one compound, noradrenaline, in the training set has a hat value higher than the warning h^* value of 1.0 in the amine system and 0.9 in the zwitterionic system, and thus is regarded as a structural outlier. This compound with a small residual belongs to the

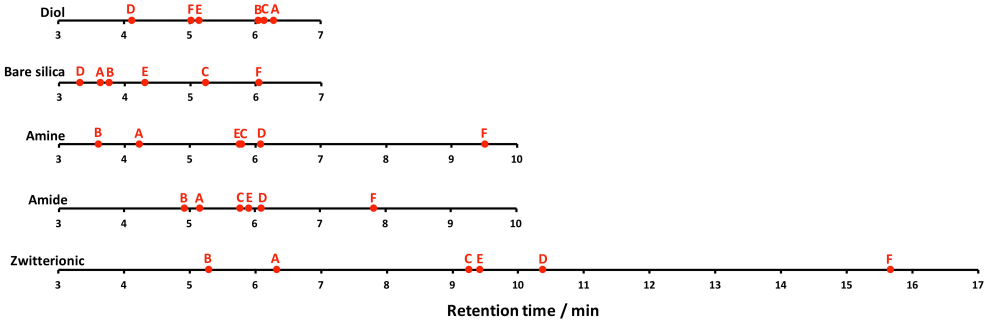

Figure 3.7. Predicted retention times (red circles) of optimised GA-PLS models for an external validation set of β-adrenergic agonists: A, 3-methylphenylamine; B, 2-phenylethylamine; C, terbutaline; D, metanephrine; E, tyramine; F, octopamine; on five HILIC stationary phases.

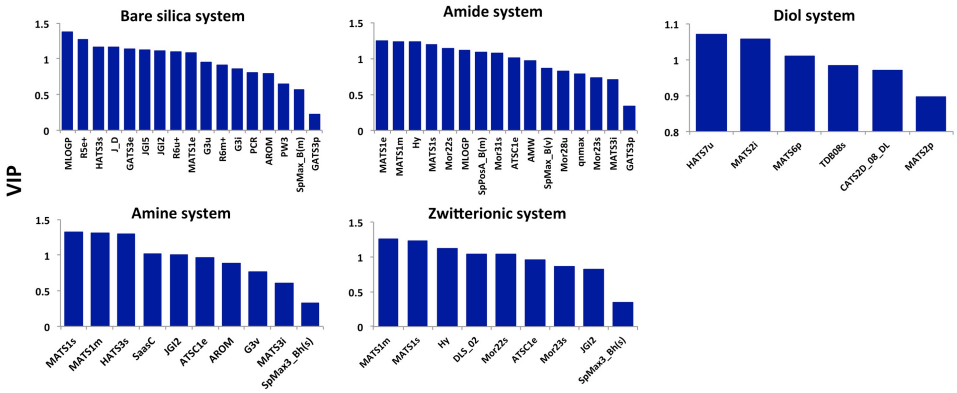

Figure 3.8. PLS variable importance to projection for the optimised GA-PLS models generated.

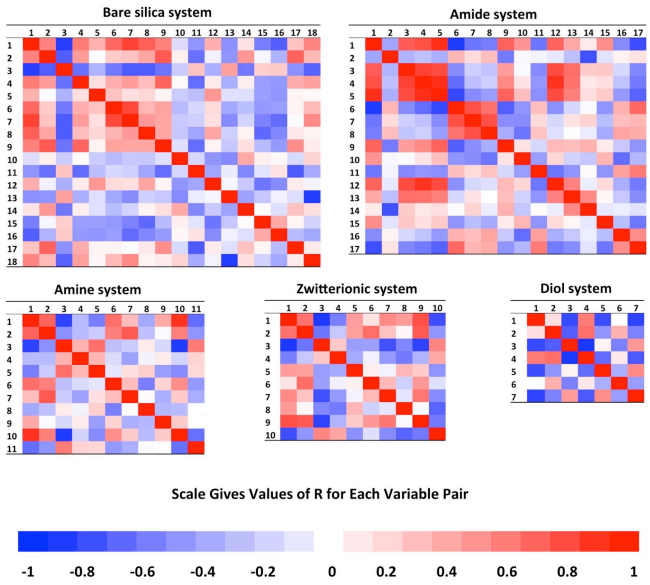

Figure 3.9. Correlation between the independent variables (descriptors) and the response variable (retention time, variable no. 1) retained in the optimised GA-PLS models over five different HILIC systems. Numbering of descriptors: 2, MATS2p; 3, MATS6p; 4, MATS2i; 5, TDB08s; 6, HATS7u; 7, CATS2D_08_DL for diol system; numbering of descriptors: 2, JGI2; 3, MLOGP; 4, JGI5; 5, G3i; 6, R6u+; 7, R5e+; 8, HATS3s; 9, R6m+; 10, GATS3p; 11, AROM; 12, G3u; 13, PCR; 14, SpMax_B(m); 15, MATS1e; 16, GATS3e; 17, PW3; 18, J_D for bare silica system; numbering of descriptors: 2, ATSC1e; 3, MATS1m; 4, MATS3i; 5, MATS1s; 6, JGI2; 7, SpMax3_Bh(s); 8, AROM; 9, G3v; 10, HATS3s; 11, SaasC for amine system; numbering of descriptors: 2, MATS3i; 3, ATSC1e; 4, SpPosA_B(m); 5, Hy; 6, MATS1m; 7, MATS1e; 8, MATS1s; 9, Mor22s; 10, Mor28u; 11, Mor31s; 12, AMW; 13, SpMax_B(v); 14, GATS3p; 15, Mor23s; 16, qnmax; 17, MLOGP for amide system and numbering of descriptors: 2, ATSC1e; 3, MATS1m; 4, GATS3e; 5, JGI2; 6, SpMax3_Bh(s); 7, Mor22s; 8, Mor23s; 9, Hy; 10, DLS_02 for zwitterionic system.

training set, so it is a good leverage compound [37]. All other compounds in the training and test sets over all HILIC systems have a hat value lower than the warning h^* value. There are a few response outliers, which are very close to the critical response value (3σ) and have small residuals.

3.3.3 Potential insights into the HILIC retention mechanism

By examination of the descriptors appearing in the proposed models for each stationary phase, some insight can be gained into the factors that influence the retention of β-adrenergic agonists on these HILIC systems. Many of the descriptors utilised are auto-correlation descriptors, both 2D and 3D. These descriptors are weighted by ionization potential, electronegativity, polarisability, mass, and I-state. In addition topological descriptors, property descriptors, and detailed 3D descriptors are also built into the final models. The molecular descriptors used and their contributions in the final QSRR models are shown in Table 3.4. The definition of each molecular descriptor is available in Table 3.5.

3D Descriptors used in the final models. 3DMoRSE (3D-Molecule Representation of Structures based on Electron diffraction) descriptors provide information derived from the weighted three-dimensional atomic coordinates by using the same transformation used in electron diffraction to prepare theoretical scattering curves [38]. For the amide system the descriptors used were Mor22s, Mor28u, Mor31s and Mor23s and for the zwitterionic system Mor22s and Mor23s (Table 3.5). These descriptors with high VIP values (Figure 3.8) are either unweighted or weighted by I-state and may show the importance of the electrostatic interaction between β-adrenergic agonists and HILIC stationary phases.

GETAWAY (Geometry, Topology, and Atom-Weights AssemblY) descriptors are geometrical descriptors that capture information on the

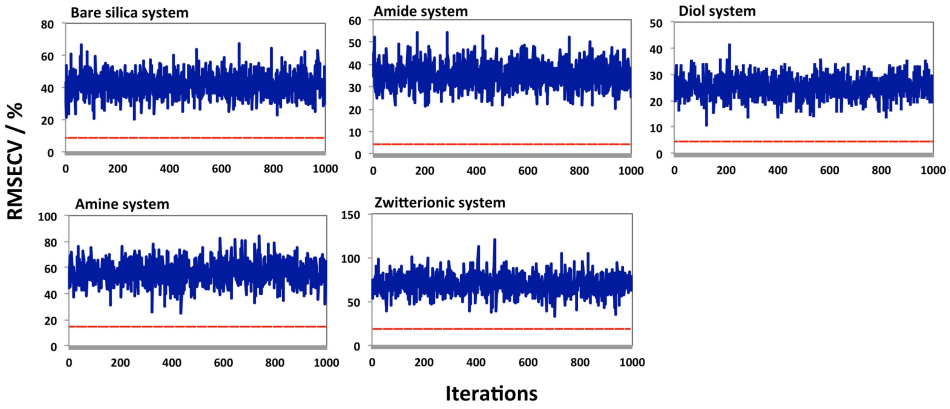

Figure 3.10. Y-randomization plot. Y-randomized data (royal blue lines), and the actual data (red line).

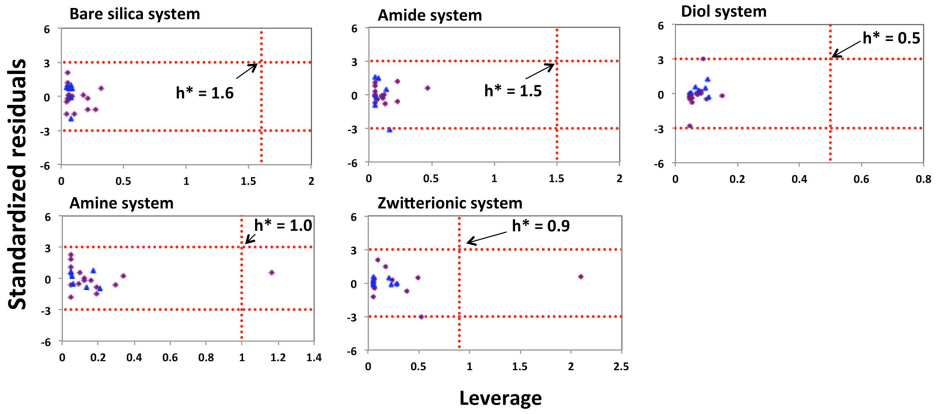

Figure 3.11. Williams plots for the optimised GA-PLS models with ±3σ, and h* (0.5, 1.6, 1.0, 1.5 and 0.9 for diol, bare silica, amine, amide and zwitterionic systems, respectively) as warning limits. Diamonds represent training set observations (n=16), and triangles represent external validation set observations (n=6).

effective position of substituents and fragments in the molecular space, and also information on molecular size and shape combined with the information on specific physicochemical atomic properties: atomic mass, atomic van der Waals volume, atomic electronegativity, and atomic polarisability [39, 40]. The GETAWAY descriptors utilised for the bare silica descriptors are either unweighted or weighted by mass, electronegativity, and I-state. The amine and diol systems also utilise GETAWAY descriptors – specifically HATS7u for the amine system and HATS3s for the diol system (Table 3.5).

The weighted holistic invariant molecular (WHIM) descriptors are calculated in such a way as to encode relevant molecular 3D information concerning molecular size, shape, symmetry, and atom distribution with respect to invariant reference frames [41, 42]. Symmetry-related WHIM descriptors in the models represent the role of the degree of the compactness of the molecules. Based on the positive sign (Table 3.4) assigned to the symmetry related WHIM descriptors G3i and G3u in the model for the bare silica system, and the G3v descriptor in the GAPLS model for the amine system it is expected that β-adrenergic agonists with higher values of these descriptors will have higher retention time values.

The geometrical descriptor AROM (aromaticity index) [43], refers to the electronic cyclic delocalization present in all aromatic species [44, 45]. As shown in table 3.4 there is a negative contribution of this descriptor to the predicted retention time values in the developed QSRR models in both amine and bare silica systems. The final 3D descriptor TDB08s appears in the final model for the diol system (Table 3.5). It belongs to a group of 3D autocorrelation descriptors that combine chemical information given by

Table 3.4. Molecular Descriptors and their Regression Coefficients (β) in the QSRR models.

Diol system		**Bare silica system**		**Amide system**		**Amine system**	
molecular descriptor	β[a]	molecular descriptor	β	molecular descriptor	β	molecular descriptor	β
MATS2p	-0.08	JGI2	0.13	MATS3i	-0.09	ATSC1e	0.40
MATS6p	-0.14	MLOGP	-0.16	ATSC1e	0.12	MATS1m	-0.54
MATS2i	0.14	JGI5	0.13	SpPosA_B(m)	0.13	MATS3i	-0.25
TDB08s	0.04	G3i	0.10	Hy	0.15	MATS1s	-0.54
HATS7u	-0.07	R6u+	0.13	MATS1m	-0.15	JGI2	0.41
CATS2D_08_DL	-0.13	R5e+	0.15	MATS1e	-0.15	SpMax3_Bh(s)	0.13
Zwitterionic system		HATS3s	0.14	MATS1s	-0.15	AROM	-0.36
molecular descriptor	β	R6m+	0.11	Mor22s	0.14	G3v	0.32
ATSC1e	0.72	GATS3p	0.03	Mor28u	0.10	HATS3s	0.53
MATS1m	-0.94	AROM	-0.09	Mor31s	-0.13	SaasC	-0.42
MATS1s	-0.92	G3u	0.11	AMW	0.12		
JGI2	0.61	PCR	-0.10	SpMax_B(v)	0.11		
SpMax3_Bh(s)	0.26	SpMax_B(m)	0.07	GATS3p	0.04		
Mor22s	0.78	MATS1e	-0.13	Mor23s	0.09		
Mor23s	0.65	GATS3e	-0.13	Qnmax	-0.10		
Hy	0.84	PW3	0.08	MLOGP	1.12		
DLS_02	-0.78	J_D	0.14				

[n] is the regression coefficients obtained from autoscaled molecular descriptors.

Table 3.5. Molecular descriptors selected by optimised QSRR models over five HILIC systems

Molecular descriptor	Description	Category
Diol system		
MATS2p	Moran autocorrelation of lag 2 weighted by polarisability	2D autocorrelations
MATS6p	Moran autocorrelation of lag 6 weighted by polarisability	2D autocorrelations
MATS2i	Moran autocorrelation of lag 2 weighted by ionization potential	2D autocorrelations
TDB08s	3D Topological distance based descriptors - lag 8 weighted by I-state	3D autocorrelations
HATS7u	leverage-weighted autocorrelation of lag 7 / unweighted	GETAWAY descriptors
CATS2D_08_DL	CATS2D Donor-Lipophilic at lag 08	CATS 2D
Bare silica system		
JGI2	mean topological charge index of order 2	2D autocorrelations
MLOGP	Moriguchi octanol-water partition coeff. (logP)	Molecular properties
JGI5	mean topological charge index of order 5	2D autocorrelations
G3i	3rd component symmetry directional WHIM index / weighted by ionization potential	WHIM descriptors
R6u+	R maximal autocorrelation of lag 6 / unweighted	GETAWAY descriptors
R5e+	R maximal autocorrelation of lag 5 / weighted by Sanderson electronegativity	GETAWAY descriptors
HATS3s	leverage-weighted autocorrelation of lag 3 / weighted by I-state	GETAWAY descriptors
R6m+	R maximal autocorrelation of lag 6 / weighted by mass	GETAWAY descriptors
GATS3p	Geary autocorrelation of lag 3 weighted by polarisability	2D autocorrelations
AROM	aromaticity index	Geometrical descriptors
G3u	3rd component symmetry directional WHIM index / unweighted	WHIM descriptors

PCR	ratio of multiple path count over path coun	Walk and path counts
SpMax_B(m)	leading eigenvalue from Burden matrix weighted by mass	2D matrix-based descriptors
MATS1e	Moran autocorrelation of lag 1 weighted by Sanderson electronegativity	2D autocorrelations
GATS3e	Geary autocorrelation of lag 3 weighted by Sanderson electronegativity	2D autocorrelations
PW3	path/walk 3 - Randic shape index	Topological indices
J_D	Balaban-like index from topological distance matrix (Balaban distance connectivity index)	2D matrix-based descriptors
Amine system		
ATSC1e	Centred Broto-Moreau autocorrelation of lag 1 weighted by Sanderson electronegativity	2D autocorrelations
MATS1m	Moran autocorrelation of lag 1 weighted by mass	2D autocorrelations
MATS3i	Moran autocorrelation of lag 3 weighted by ionization potential	2D autocorrelations
MATS1s	Moran autocorrelation of lag 1 weighted by I-state	2D autocorrelations
JGI2	mean topological charge index of order 2	2D autocorrelations
SpMax3_Bh(s)	largest eigenvalue n. 3 of Burden matrix weighted by I-state	Burden eigenvalues
AROM	aromaticity index	Geometrical descriptors
G3v	3rd component symmetry directional WHIM index / weighted by van der Waals volume	WHIM descriptors
HATS3s	leverage-weighted autocorrelation of lag 3 / weighted by I-state	GETAWAY descriptors
SaasC	Sum of aasC E-states	Atom-type E-state indices
Amide system		
MATS3i	Moran autocorrelation of lag 3 weighted by ionization potential	2D autocorrelations
ATSC1e	Centred Broto-Moreau autocorrelation of lag 1 weighted by Sanderson electronegativity	2D autocorrelations

SpPosA_B(m)	normalised spectral positive sum from Burden matrix weighted by mass	2D matrix-based descriptors
Hy	hydrophilic factor	Molecular properties
MATS1m	Moran autocorrelation of lag 1 weighted by mass	2D autocorrelations
MATS1e	Moran autocorrelation of lag 1 weighted by Sanderson electronegativity	2D autocorrelations
MATS1s	Moran autocorrelation of lag 1 weighted by I-state	2D autocorrelations
Mor22s	signal 22 / weighted by I-state	3D-MoRSE descriptors
Mor28u	signal 28 / unweighted	3D-MoRSE descriptors
Mor31s	signal 31 / weighted by I-state	3D-MoRSE descriptors
AMW	average molecular weight	Constitutional indices
SpMax_B(v)	leading eigenvalue from Burden matrix weighted by van der Waals volume	2D matrix-based descriptors
GATS3p	Geary autocorrelation of lag 3 weighted by polarisability	2D autocorrelations
Mor23s	signal 23 / weighted by I-state	3D-MoRSE descriptors
qnmax	maximum negative charge	Charge descriptors
MLOGP	Moriguchi octanol-water partition coeff. (logP)	Molecular properties
Zwitterionic system		
ATSC1e	Centred Broto-Moreau autocorrelation of lag 1 weighted by Sanderson electronegativity	2D autocorrelations
MATS1m	Moran autocorrelation of lag 1 weighted by mass	2D autocorrelations
MATS1s	Moran autocorrelation of lag 1 weighted by I-state	2D autocorrelations
JGI2	mean topological charge index of order 2	2D autocorrelations
SpMax3_Bh(s)	largest eigenvalue n. 3 of Burden matrix weighted by I-state	Burden eigenvalues
Mor22s	signal 22 / weighted by I-state	3D-MoRSE descriptors
Mor23s	signal 23 / weighted by I-state	3D-MoRSE descriptors
Hy	hydrophilic factor	Molecular properties
DLS_02	modified drug-like score from Oprea et al. (6 rules)	Drug-like indices

property values in specified molecule regions, and structural information [46]. This descriptor is a topological distance descriptor weighted by the I-state and therefore relates to the local electronic characteristics of the studied compounds [47].

2D Descriptors used in the final models. 2D autocorrelation descriptors provide information about the distribution of a selected physicochemical property along a topological map of a molecular structure [12]. These properties are atomic masses, polarisabilities, charges, and electronegativities. The appearance of the atomic mass, electronegativity-weighted, polarisability-weighted and charge-related 2D autocorrelation descriptors in the developed models (Table 3.5) represents the role of atomic size and electronic properties in the retention behaviour of β-adrenergic agonists in HILIC systems.

As shown in Table 3.4 and Table 3.5, a large number of different 2D autocorrelation descriptors were utilised in the models for the different systems. The index for the mean topological charge JGI2 was selected for the model for the bare silica system, the amine system and the zwitterionic system while JGI5 was only selected for the bare silica system (Table 3.5). GATS3p, weighted by polarisability, was selected for use in the bare silica and amide systems but with very low importance (Table 3.4, Table 3.5, Figure 3.8). GATS3e, weighted by electronegativity, had a much higher importance but was only utilised in the model for the bare silica column (Table 3.4, Table 3.5).

A Moran autocorrelation index weighted by electronegativity (MATS1e) was utilised in the models for both the bare silica and the amide systems, while similar indices weighted by mass and I-state (MATS1m and MATS1s) were utilised in the models for the amine, amide and zwitterionic systems (Table 3.4, Table 3.5). Two different Moran autocorrelation indices

weighted by ionization potential (MATS2i and MATS3i) were utilised for the diol, amine, and amide system models but indices weighted by polarisability were only utilised in the model for the diol system (Table 3.4, Table 3.5). The inclusion of the Moran autocorrelation indices in all models with the different weightings used suggest slightly different mechanisms at play in the different column types.

Similar 2D autocorrelation indices are the Geary and the Centred Broto-Moreau autocorrelation. The Centred Broto-Moreau autocorrelation index is included in the models for the amine, amide and zwitterionic systems (ATSC1e weighted by electronegativity), while Geary indices (GATS3p and GATS3e) are included in the bare silica and amide columns (Table 3.4, Table 3.5). GATS3e weighed by electronegativity is of far higher importance (by VIP) and only present in the bare silica system model (Table 3.4, Table 3.5, Figure 3.8).

The 2D matrix-based descriptors are topological indices calculated from different graph-theoretical matrices derived from the H-depleted molecular graph of molecules [48]. They encode information about atom connectivity [12]. It is evident from the sign of regression coefficient of the participating descriptors SpMax_B(m) and J_D in the model for the bare silica system, and SpMax_B(v) and SpPosA_B(m) in the model for the amide system that the descriptors from this class have contributed positively to the retention time (Table 3.4).

SpMax3_Bh(s) belongs to the set of Burden eigenvalues. These descriptors are calculated from the connectivity matrix using Burden [49] procedures and encode information about the structure topology, including the bond order. The VIP value of the SpMax3_Bh(s) descriptor indicates that this descriptor has the lowest importance in the developed QSRR models in the zwitterionic and amine systems (Figure 3.8).

Descriptor PW3 incorporated in the QSRR model generated for the bare silica system (Table 3.4, Table 3.5) belongs to the category of topological indices [50, 51]. This descriptor is measured by applying the ratios of the atomic path count over the atomic walk count and the number of non-H atoms [52].

The Chemically Advanced Template Search (CATS) 2D descriptors are topological descriptors calculated based on the distance between all possible pairs of pharmacophore elements in a molecule [53]. The pharmacophore elements are anions, cations, hydrogen bond acceptors, hydrogen bond donors, and hydrophobic atoms. The CATS2D_08_DL descriptor from the CATS2D class is one of the most important factors for retention time in the diol system, and represents the hydrogen bond effect with the present of heteroatom, O, and the lipophilicity interaction (Table 3.4, Table 3.5). The negative regression of the CATS2D_08_DL descriptor (Table 3.4) indicates that a higher positive value is correlated with a decreased retention time. This result is not surprising because it is well known that hydrogen bonding and hydrophobic partitioning describe the retention mechanism of β-adrenergic agonists and related compounds in HILIC systems [5, 8].

Atom-type E-state indices, SaasC, selected for the amine system model, are proposed as molecular descriptors encoding topological and electronic information related to particular atom types in the molecule [54]. The high VIP value (Figure 3.8) of this descriptor presents the importance of electrostatic interactions between β-adrenergic agonists and amine stationary phases.

Other Descriptors used in the final models. Important molecular properties incorporated in the QSRR models are reflected by the structural descriptors MLOGP (Moriguchi octanol-water partition coeff. (logP)) and Hy (hydrophilic factor). The MLOGP is a measure of the lipophilicity of the

molecule, which is estimated using the Moriguchi method based on he structure-logP relationship of 1230 organic molecules [55]. LogP is the main factor influencing the retention behaviour of β-adrenergic agonists in the bare silica system with a VIP value of 1.38 (Figure 3.8). This descriptor also is incorporated in the equation constructed for amide system as one of the highest contributing descriptors (Table 3.4, Table 3.5), and has been reported to contribute to HILIC mechanism previously [18, 19, 21]. As shown in Table 3.4, LogP has a negative coefficient in both final QSRR models. The lower the logP values, the greater the retention times. This is consistent with the fact that compounds with high logP values have low hydrophilicity and therefore low retention in the water-enriched layer in HILIC mode [18]. Descriptor Hy has been found to be one of the most important parameters in QSRR models developed for the zwitterionic and amide systems with a positive contribution to the retention time (Table 3.4, Table 3.5). This well-known descriptor in HILIC mechanism reflects the ability of a solute to participate in hydrophilic partitioning between a water-enriched layer immobilized at the stationary phase surface and the bulk organic-rich mobile phase [1].

The constitutional descriptor, AMW contributed to the model prediction in the amide system and reflects the chemical composition of a compound independent from its molecular geometry or atom connectivity. AMW represents average molecular weight, which contributes to dispersive intermolecular interactions [56] and inclusion of this descriptor has a positive effect on predicted retention time with a VIP value of 0.97 (Table 3.4, Table 3.5, Figure 3.8).

Descriptor qnmax (maximum negative charge) belongs to the category of charge descriptors and encodes features responsible for electrostatic and ionic interactions between molecules [57]. The descriptor qnmax has

negative values and a negative coefficient (Table 3.4) in the QSRR model derived for the amide system. Therefore a lower value of this descriptor increases the predicted retention time.

Drug-like indices are calculated based on rules defined for drug elements: H-bond donors, H-bond acceptors, molecular weight, MLOGP, rotatable bond number and polar surface area [58]. The descriptor DLS_02 of this class negatively contributes to the final model derived for the zwitterionic system (Table 3.4, Table 3.5) with a VIP value more than 1 (Figure 3.8). Thus, it can be concluded that the β-adrenergic agonist having lower value of this descriptor will be retained in the stationary phase longer.

Walk and path counts, PCR, in the final QSRR model calculated for the bare silica system (Table 3.4, Table 3.5) characterises structural aspects of a molecule by investigating the bond paths between atoms in the molecule [51].

It can be seen from the above discussion that the mechanisms that are likely to govern the retention behaviour of β-adrenergic agonists in HILIC systems are complex and that a complex model such as that built through the optimised GA-PLS method is necessary to predict retention.

3.3.4 Conclusions

Our QSRR methodology can be used to predict the retention of β-adrenergic agonists with a high degree of accuracy and precision over five HILIC systems. The models are easily derived from a small set of β-adrenergic agonists and their experimentally measured retention times. Quick theoretical calculations provide models that allow a researcher to readily distinguish suitable stationary phases for the study under development.

One of the novelties of the method described herein is the application of GA to select an optimised descriptor set for the QSRR models. The employment of optimised GA-PLS allowed us to identify the most relevant descriptors before obtaining the model, and to propose robust QSRR models, despite the high complexity of retention mechanisms in HILIC systems. In addition, models from GA-PLS based on the optimised descriptor sets outperformed those from GA-PLS and full PLS calculations, therefore we suggest the use of the former in the development of HILIC studies. Furthermore, the feature selection approach represented is shown to be a powerful tool to capture the general aspects of interactions for different HILIC systems.

3.4 References

[1] A.J. Alpert, Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds, J. Chromatogr. A, 499 (1990) 177-196.

[2] B. Buszewski, S. Noga, Hydrophilic interaction liquid chromatography (HILIC)-a powerful separation technique, Anal. Bioanal. Chem., 402 (2012) 231-247.

[3] P. Hemström, K. Irgum, Hydrophilic interaction chromatography, J. Sep. Sci., 29 (2006) 1784-1821.

[4] Z. Aturki, G. D'Orazio, A. Rocco, K. Si-Ahmed, S. Fanali, Investigation of polar stationary phases for the separation of sympathomimetic drugs with nano-liquid chromatography in hydrophilic interaction liquid chromatography mode, Anal. Chim. Acta, 685 (2011) 103-110.

[5] R.I. Chirita, C. West, A.L. Finaru, C. Elfakir, Approach to hydrophilic interaction chromatography column selection: application to neurotransmitters analysis, J. Chromatogr. A, 1217 (2010) 3091-3104.

[6] A. Kumar, J.P. Hart, D.V. McCalley, Determination of catecholamines in urine using hydrophilic interaction chromatography with electrochemical detection, J. Chromatogr. A, 1218 (2011) 3854-3861.
[7] F. Gosetti, E. Mazzucco, M.C. Gennaro, E. Marengo, Simultaneous determination of sixteen underivatized biogenic amines in human urine by HPLC-MS/MS, Anal. Bioanal. Chem., 405 (2013) 907-916.
[8] S. Tufi, M. Lamoree, J. de Boer, P. Leonards, Simultaneous analysis of multiple neurotransmitters by hydrophilic interaction liquid chromatography coupled to tandem mass spectrometry, J. Chromatogr. A, 1395 (2015) 79-87.
[9] A. Periat, B. Debrus, S. Rudaz, D. Guillarme, Screening of the most relevant parameters for method development in ultra-high performance hydrophilic interaction chromatography, J. Chromatogr. A, 1282 (2013) 72-83.
[10] G. Schuster, W. Lindner, Comparative characterisation of hydrophilic interaction liquid chromatography columns by linear solvation energy relationships, J. Chromatogr. A, 1273 (2013) 73-94.
[11] R. Kaliszan, QSRR: quantitative structure-(chromatographic) retention relationships, Chem. Rev., 107 (2007) 3212-3246.
[12] R. Todeschini, V. Consonni, Molecular Descriptors for Chemoinformatics, Wiley-WCH, Weinheim, 2009.
[13] R. Put, Y. Vander Heyden, Review on modelling aspects in reversed-phase liquid chromatographic quantitative structure-retention relationships, Anal. Chim. Acta, 602 (2007) 164-172.
[14] M. Goodarzi, R. Jensen, Y. Vander Heyden, QSRR modelling for diverse drugs using different feature selection methods coupled with linear and nonlinear regressions, J. Chromatogr. B, 910 (2012) 84-94.
[15] M. Talebi, G. Schuster, R.A. Shellie, R. Szucs, P.R. Haddad, Performance comparison of partial least squares-related variable selection

methods for quantitative structure retention relationships modelling of retention times in reversed-phase liquid chromatography, J. Chromatogr. A, 1424 (2015) 69-76.
[16] P. Zuvela, J.J. Liu, K. Macur, T. Baczek, Molecular descriptor subset selection in theoretical peptide quantitative structure-retention relationship model development using nature-inspired optimisation algorithms, Anal. Chem., 87 (2015) 9876-9883.
[17] F. Tian, L. Yang, F. Lv, P. Zhou, Predicting liquid chromatographic retention times of peptides from the Drosophila melanogaster proteome by machine learning approaches, Anal. Chim. Acta, 644 (2009) 10-16.
[18] K. Gorynski, B. Bojko, A. Nowaczyk, A. Bucinski, J. Pawliszyn, R. Kaliszan, Quantitative structure-retention relationships models for prediction of high performance liquid chromatography retention time of small molecules: endogenous metabolites and banned compounds, Anal. Chim. Acta, 797 (2013) 13-19.
[19] D.J. Creek, A. Jankevics, R. Breitling, D.G. Watson, M.P. Barrett, K.E. Burgess, Toward global metabolomics analysis with hydrophilic interaction liquid chromatography-mass spectrometry: improved metabolite identification by retention time prediction, Anal. Chem., 83 (2011) 8703-8710.
[20] K. Jinno, N.S. Quiming, N.L. Denola, Y. Saito, Modelling of retention of adrenoreceptor agonists and antagonists on polar stationary phases in hydrophilic interaction chromatography: a review, Anal. Bioanal. Chem., 393 (2009) 137-153.
[21] M. Cao, K. Fraser, J. Huege, T. Featonby, S. Rasmussen, C. Jones, Predicting retention time in hydrophilic interaction liquid chromatography mass spectrometry and its use for peak annotation in metabolomics, Metabolomics, 11 (2015) 696-706.

[22] B.J. Johnson, S.B. Smith, K.Y. Chung, Historical Overview of the Effect of beta-Adrenergic Agonists on Beef Cattle Production, Asian-Australas. J. Anim. Sci., 27 (2014) 757-766.
[23] S.B. Liggett, S.D. Shah, P.E. Cryer, Increased fat and skeletal muscle beta-adrenergic receptors but unaltered metabolic and hemodynamic sensitivity to epinephrine in vivo in experimental human thyrotoxicosis, J. Clin. Invest., 83 (1989) 803-809.
[24] H. Kuiper, M. Noordam, M. van Dooren-Flipsen, R. Schilt, A. Roos, Illegal use of beta-adrenergic agonists: European Community, J. Anim. Sci., 76 (1998) 195-207.
[25] L. Beucher, G. Dervilly-Pinel, S. Prevost, F. Monteau, B. Le Bizec, Determination of a large set of beta-adrenergic agonists in animal matrices based on ion mobility and mass separations, Anal. Chem., 87 (2015) 9234-9242.
[26] X.D. Du, Y.L. Wu, H.J. Yang, T. Yang, Simultaneous determination of 10 beta2-agonists in swine urine using liquid chromatography-tandem mass spectrometry and multi-walled carbon nanotubes as a reversed dispersive solid phase extraction sorbent, J. Chromatogr. A, 1260 (2012) 25-32.
[27] M. Leporati, M. Bergoglio, P. Capra, E. Bozzetta, M.C. Abete, M. Vincenti, Development, validation and application to real samples of a multiresidue LC-MS/MS method for determination of beta2 -agonists and anabolic steroids in bovine hair, J. Mass Spectrom., 49 (2014) 936-946.
[28] G.J. Zhang, B.H. Fang, Y.H. Liu, X.F. Wang, L.X. Xu, Y.P. Zhang, L.M. He, Development of a multi-residue method for fast screening and confirmation of 20 prohibited veterinary drugs in feedstuffs by liquid chromatography tandem mass spectrometry, J. Chromatogr. B, 936 (2013) 10-17.
[29] N.S. Quiming, N.L. Denola, I. Ueta, Y. Saito, S. Tatematsu, K. Jinno, Retention prediction of adrenoreceptor agonists and antagonists on a diol

column in hydrophilic interaction chromatography, Anal. Chim. Acta, 598 (2007) 41-50.
[30] N.S. Quiming, N.L. Denola, S.R. Samsuri, Y. Saito, K. Jinno, Development of retention prediction models for adrenoreceptor agonists and antagonists on a polyvinyl alcohol-bonded stationary phase in hydrophilic interaction chromatography, J. Sep. Sci., 31 (2008) 1537-1549.
[31] N.S. Quiming, N.L. Denola, Y. Saito, K. Jinno, Retention prediction of adrenoreceptor agonists and antagonists on unmodified silica phase in hydrophilic interaction chromatography, Anal. Bioanal. Chem., 388 (2007) 1693-1706.
[32] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemom. Intell. Lab. Syst., 58 (2001) 109-130.
[33] J. Ghasemi, S. Saaidpour, QSRR Prediction of the Chromatographic Retention Behaviour of Painkiller Drugs, J. Chromatogr. Sci., 47 (2009) 156-163.
[34] T. Mehmood, K.H. Liland, L. Snipen, S. Sæbø, A review of variable selection methods in Partial Least Squares Regression, Chemom. Intell. Lab. Syst., 118 (2012) 62-69.
[35] I.-G. Chong, C.-H. Jun, Performance of some variable selection methods when multicollinearity is present, Chemom. Intell. Lab. Syst., 78 (2005) 103-112.
[36] J. Huang, X. Fan, Why QSAR fails: an empirical evaluation using conventional computational approach, Mol. Pharmaceutics, 8 (2011) 600-608.
[37] P. Gramatica, Principles of QSAR models validation: internal and external, QSAR Comb. Sci., 26 (2007) 694-701.
[38] O. Devinyak, D. Havrylyuk, R. Lesyk, 3D-MoRSE descriptors explained, J. Mol. Graph. Model., 54 (2014) 194-203.

[39] V. Consonni, R. Todeschini, M. Pavan, Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors. 1. Theory of the Novel 3D Molecular Descriptors, J. Chem. Inf. Model., 42 (2002) 682-692.
[40] V. Consonni, R. Todeschini, M. Pavan, P. Gramatica, Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors. 2. Application of the Novel 3D Molecular Descriptors to QSAR/QSPR Studies, J. Chem. Inf. Model., 42 (2002) 693-705.
[41] R. Todeschini, P. Gramatica, SD-modelling and Prediction by WHIM Descriptors. Part 5. Theory Development and Chemical Meaning of WHIM Descriptors, Quant. Struct.-Act. Relat., 16 (1997) 113-119.
[42] R. Todeschini, New 3D molecular descriptors: the WHIM theory and QSAR applications, Perspect. Drug Discovery Des., 9/11 (1998) 355-380.
[43] A.R. Katritzky, E.V. Gordeeva, Traditional topological indexes vs electronic, geometrical, and combined molecular descriptors in QSAR/QSPR research, J. Chem. Inf. Model., 33 (1993) 835-857.
[44] T.M. Krygowski, K. Ejsmont, B.T. Stepien, M.K. Cyranski, J. Poater, M. Sola, Relation between the substituent effect and aromaticity, J. Org. Chem., 69 (2004) 6634-6640.
[45] F. Feixas, E. Matito, J. Poater, M. Sola, On the performance of some aromaticity indices: a critical assessment using a test set, J. Comput. Chem., 29 (2008) 1543-1554.
[46] R. Todeschini, V. Consonni, Descriptors from Molecular Geometry, (2003) 1004-1033.
[47] C.T. Klein, D. Kaiser, G. Ecker, Topological distance based 3D descriptors for use in QSAR and diversity analysis, J. Chem. Inf. Comput. Sci., 44 (2004) 200-209.

[48] V. Consonni, R. Todeschini, Multivariate Analysis of Molecular Descriptors, (2012) 111-147.
[49] F.R. Burden, Molecular identification number for substructure searches, J. Chem. Inf. Model., 29 (1989) 225-227.
[50] R. Todeschini, R. Cazar, E. Collina, The chemical meaning of topological indices, Chemom. Intell. Lab. Syst., 15 (1992) 51-59.
[51] M. Randic, J. Zupan, On Interpretation of Well-Known Topological Indices, J. Chem. Inf. Model., 41 (2001) 550-560.
[52] M. Randic, Novel Shape Descriptors for Molecular Graphs, J. Chem. Inf. Model., 41 (2001) 607-613.
[53] G. Schneider, W. Neidhart, T. Giller, G. Schmid, "Scaffold-Hopping" by Topological Pharmacophore Search: A Contribution to Virtual Screening, Angew. Chem., Int. Ed., 38 (1999) 2894-2896.
[54] Q.N. Hu, Y.Z. Liang, H. Yin, X.L. Peng, K.T. Fang, Structural interpretation of the topological index. 2. The molecular connectivity index, the Kappa index, and the atom-type E-State index, J. Chem. Inf. Comput. Sci., 44 (2004) 1193-1201.
[55] I. Moriguchi, S. Hirono, Q. Liu, I. Nakagome, Y. Matsushita, Simple Method of Calculating Octanol/Water Partition Coefficient, Chem. Pharm. Bull., 40 (1992) 127-130.
[56] R. Kaliszan, A. Kaliszan, T.A.G. Noctor, W.P. Purcell, I.W. Wainer, Mechanism of retention of benzodiazepines in affinity, reversed-phase and adsorption high-performance liquid chromatography in view of quantitative structure retention relationships, J. Chromatogr. A, 609 (1992) 69-81.
[57] J. Galvez, R. Garcia, M.T. Salabert, R. Soler, Charge Indexes. New Topological Descriptors, J. Chem. Inf. Model., 34 (1994) 520-525.
[58] T.I. Oprea, Property distribution of drug-related chemical databases, J. Comput.-Aided Mol. Des., 14 (2000) 251-264.

4 Rapid method development in hydrophilic interaction liquid chromatography for pharmaceutical analysis using a combination of quantitative structure-retention relationships and design of experiments

4.1 Introduction

The process of finding optimal separation conditions on the basis of reliable theoretical predictions is an important step in rational method development for high-performance liquid chromatography (HPLC). Accordingly, the possibility of prediction of retention and separation in HPLC in the absence of any prior experiments with a sufficient accuracy to support method development is continuously moving further into the focus of theoretical and experimental chemists [1]. Since adoption of Quality-by-Design (QbD) concepts in the pharmaceutical industry by the International Conference on Harmonisation (ICH) guidelines Q8(R2) [2], an important prerequisite to identify optimal separation conditions has been the applicability of QbD principles which allow for more robust and reliable analytical methods with fewer method failures or method transfer issues. A QbD-based treatment of the robustness of an HPLC method requires the application of the Design of Experiments (DoE) philosophy to establish a comprehensive design space, which can be further analysed to determine the experimental conditions that provide the required optimal level of performance in terms of separation of analytes [3]. DoE is a standard methodology for reversed phase liquid chromatography (RPLC) method development [4], both in academia and industry. This concept has currently been employed in the commercial optimisation software Drylab [5].

To date, researchers have highlighted the possibility of applying QbD concepts to those analytes having experimental data available in order to

determine their retention under changing chromatographic conditions [6-8]. However, there exist strategies capable of developing relationships linking chromatographic parameters and analyte properties, and to automate their use in the prediction of retention parameters of new analytes. To the best of our knowledge, quantitative structure-retention relationships (QSRR) based on theoretically generated molecular descriptors [9] have so far attracted attention in only two QbD studies [10, 11]. Such a technique, however, has been successfully applied in LC method development [12, 13]. Thus, an approach that combines both DoE and QSRR methodologies might be of help to overcome the limitations of QbD techniques that are based solely on DoE modelling equations and can only predict the retention of known analytes under new conditions and not the retention of new analytes.

The retention time prediction accuracy of a QSRR model may sometimes not be sufficient to support detailed method development [9]. However, QSRR models built using a method that clusters compounds according to structural similarity may overcome this limitation [10, 14, 15]. In this study, a novel compound-classification based QSRR modelling strategy is presented applying the concept of molecular similarity [16].

Hydrophilic interaction liquid chromatography (HILIC) is quickly becoming popular in the pharmaceutical industry due to its suitability for the separation of polar molecules [17]. A mixture of pharmaceuticals analysed in the HILIC mode was chosen as test analytes for this study (see Section 4.2 for details). First, DoE principles were applied to derive a response equation for prediction of retention times over a design space of mobile phase compositions selected using a central composite design experimental plan. The accuracy of prediction was evaluated for mobile phase compositions not used in the derivation of the model. Next, it is shown that cluster-based QSRR modelling (where analytes are clustered according to

structural similarity) can be used to predict HILIC retention times for new analytes, based only on their chemical structures, and that these predicted retention times can then be used in DoE modelling. This combination of cluster-based QSRR and DoE allows the prediction of retention for new analytes and new mobile phase compositions not used in the derivation of the models. Finally, the QSRR-DoE-computed retention times of the test probes and subsequently calculated separation selectivity were used to make a prediction about the robust areas of the design space. Experiments performed to evaluate the validity of the QbD predictions showed convincing agreement between experiment and theory.

This chapter demonstrates that discovery of optimal conditions for separation of new analytes not used in any of the modelling steps can be accelerated by integration of DoE and QSRR methods.

4.2 Method

4.2.1 Data set

A data set composed of 50 pharmaceutical compounds using a HILIC amide stationary phase was used. This data involved 1, 2'-deoxyadenosine; 2, 2',3'-dideoxyadenosine; 3, 2'-deoxyguanosine; 4, 3'-deoxyguanosine; 5, 5-methyluridine; 6, adenosine; 7, guanosine; 8, inosine; 9, thymidine; 10, uridine; 11, 3'-deoxythymidine; 12, 2'-deoxyuridine; 13, 2'-deoxyinosine; 14, acyclovir; 15, salicylic acid; 16, 5-methylsalicylic acid; 17, 4-hydroxybenzoic acid; 18, 3-hydroxybenzoic acid; 19, 2,4-dihydroxybenzoic acid; 20, 2,5-dihydroxybenzoic acid; 21, 2,3-dihydroxybenzoic acid; 22, 3,5-dihydroxybenzoic acid; 23, benzoic acid; 24, 4-aminosalicylic acid; 25, 3-amino-4-hydroxybenzoic acid; 26, 4-aminobenzoic acid; 27, 3-aminobenzoic acid; 28, p-toluic acid; 29, vanillic acid; 30, syringic acid; 31, 2-methoxybenzoic acid; 32, 3-methoxytyramine; 33, adrenaline; 34, dopamine; 35, isoproterenol; 36, metaproterenol; 37, N-methylephedrine;

38, noradrenaline; 39, norfenefrine; 40, phenylephrine; 41, ritudrine; 42, salbutamol; 43, synephrine; 44, tyramine; 45, normetanephrine; 46, fenotrole; 47, terbutaline; 48, octopamine; 49, methoxamine; 50, isoxuprine.

Experimental retention data collected under the 17 chromatographic conditions of the studied central composite design using the amide column were utilised for this study. The levels studied for the selected critical chromatographic parameters and a diagram of a central composite design with 17 independent trials are shown in Table 2.1, Table 2.2 and Figure 2.1. More information about the collected data is presented in Chapter 2.

4.2.2 Compound classification

In the classification analysis step, Exclusion Spheres [18] cluster analysis was modified (see below) and carried out using ChemAxon hashed fingerprint [19] descriptors and the Tanimoto similarity coefficient [20].

The key points in the Exclusion Spheres clustering method were identification of the cluster centroid, which is the molecule with the largest number of similar neighbours and specification of the Tanimoto threshold needed for the algorithm execution. However, this method tends, in some cases, to create heterogeneity within a cluster, since it considers the Tanimoto value of similarity of each molecule to just the centroid molecule. In this chapter, a new clustering algorithm applying the exclusion spheres principles is introduced in which homogeneity within a cluster reflects the desired Tanimoto index of similarity between all members in the cluster.

The in-house algorithm considered the following steps:

Step 1: *Generation of a matrix of pairwise similarities*. To generate a matrix of pairwise similarities, each compound's pairwise Tanimoto similarity index to each other compound in the dataset was calculated using

JChem for Excel (ChemAxon, Budapest, Hungary). The initial dataset includes more than 500 compounds collected from HILIC literature.

Step 2: *Identifying the compounds with the largest number of neighbours*. To identify such a molecule, the number of neighbours for each molecule in the dataset was calculated using the Tanimoto similarity threshold value of 0.5 (defined as a clustering parameter). The dataset was then sorted in descending order, so that the compound with the largest number of neighbours was placed at the top of the list.

Step 3. *Cluster Algorithm*. This step started with the first compound in the sorted list from step 2 as the initial cluster member. The set of its neighbours was sorted in descending order based on their pairwise Tanimoto values to the first member of the cluster, so that the compound with the highest Tanimoto value was labeled as the nearest neighbour. The next step was to count the nearest neighbour as another member of the cluster, so that now the cluster had two members. Then, pairwise similarities of the second nearest neighbour to each member of the cluster were computed and if all pairwise Tanimoto values were above, or equal to, the threshold value, it became another member of that cluster; otherwise it was discarded. This step was iterated until all other neighbours had been either selected or excluded. In the next step, this process was repeated for the excluded neighbours to keep those which were similar to at least 80% (defined by user) of all the members in that cluster based on the threshold value. Once the first cluster was formed, the members of that cluster were removed from any further comparisons and the same process was repeated for all other remaining molecules. The overall approach is illustrated in Figure 4.1 and Figure 4.2 shows the Tanimoto similarity matrix for the final clustered dataset. This compound classification step is followed by

generation of a QSRR model on each obtained cluster. Details of the model generation are presented in Chapter 2.

Following data collection (experimental retention data are shown in Table 4.1), DoE models were constructed by applying multiple linear regression [21], which is the most common regression technique used in DoE studies [6, 22, 23]. A test condition, separated from the training conditions (which are listed in Table 2.2) was used to validate the final models for their predictive ability by calculating RMSEP% as defined in Chapter 2.

4.2.3 QbD methodology

The purpose of this study was to develop a HILIC method in accordance with QbD principles for the separation of the studied pharmaceutical targets, quickly providing robust HILIC mobile phase regions that will give flexibility in routine work. The QbD workflow followed is represented schematically in Figure 1.2.

The initial step consisted of the selection of a critical quality attribute (CQA) [2], which was defined as the best separation of all targets. Selectivity factor values (α) were calculated using $\alpha = k_a/k_b$, where k_a and k_b are the retention factors of analyte a and analyte b. The critical value was defined as $\alpha \geq 1.15$ in order to identify the set of experimental conditions with acceptable method performance.

One of the essential elements of QbD is to apply the DoE approach, discussed earlier. The collected data using DoE principles was then used as an input for the next step.

A QSRR model was generated and used to predict the retention time and consequently the selectivity factor (α) between new target analytes not considered in previous modelling steps. The predicted GA-PLS data were

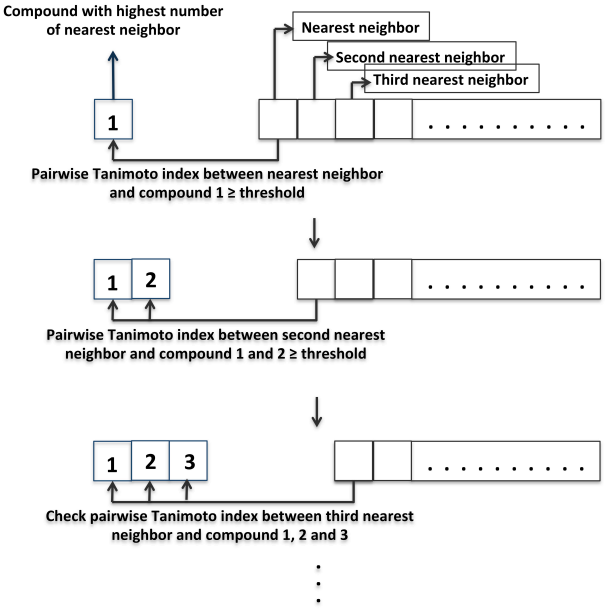

Figure 4.1. Scheme of in-house clustering algorithm.

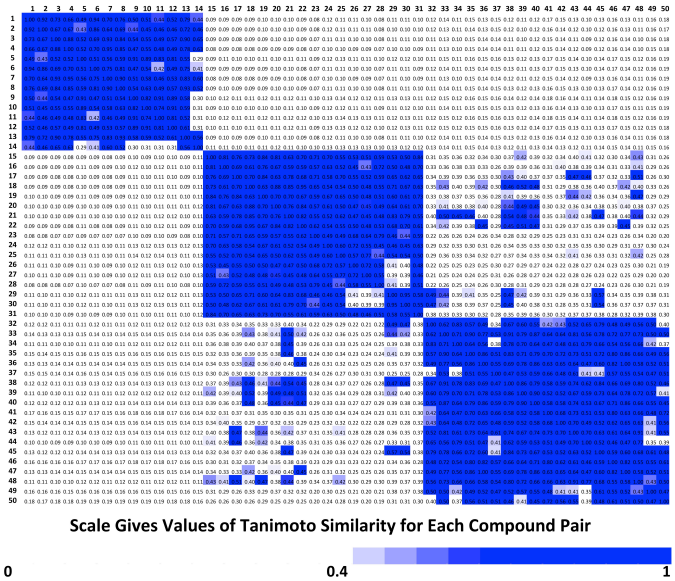

Figure 4.2. Tanimoto similarity matrix for the datasets studied. Pairwise Tanimoto values are shown, with all values >0.4 being shaded according to the color code at the bottom of the Figure. The darkest shading represents Tanimoto values >0.5. Numbering of compounds is available in Section 4.2.1.

Table 4.1. Experimentally obtained retention times for each operating condition of the central composite design.

nr.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1.92	5.98	2.39	6.55	2.17	6.16	2.47	7.19	2.33	6.64	2.78	3.21	2.63	3.15	3.07	3.04	3.07
2	1.93	6.40	2.44	7.20	2.21	6.66	2.54	8.45	2.36	7.20	2.86	3.33	3.13	3.24	3.15	3.15	3.15
3	1.91	6.75	2.33	6.83	2.15	6.83	2.41	7.42	2.30	6.85	2.79	3.13	3.00	3.07	3.01	2.98	3.03
4	1.91	5.88	2.32	6.25	2.15	5.98	2.40	6.70	2.28	6.22	2.74	3.10	2.96	3.04	2.97	2.95	2.99
5	1.93	6.49	2.39	6.99	2.20	6.68	2.48	7.62	2.35	6.99	2.86	3.29	3.12	3.22	3.13	3.11	3.14
6	1.94	6.41	2.37	6.86	2.18	6.62	2.46	7.46	2.33	6.88	2.84	3.24	3.08	3.18	3.09	3.07	3.11
7	2.02	7.64	2.57	8.38	2.28	8.16	2.61	9.54	2.48	8.60	3.09	3.52	3.43	3.49	3.41	3.39	3.42
8	1.81	3.98	2.25	4.34	2.04	3.89	2.31	4.40	2.19	4.26	2.43	2.79	2.66	2.71	2.65	2.64	2.68
9	1.99	7.58	2.48	8.12	2.25	7.99	2.57	9.06	2.41	8.23	3.02	3.51	3.31	3.43	3.34	3.31	3.35
10	1.94	6.06	2.41	6.75	2.18	6.24	2.50	7.28	2.35	6.73	2.82	3.30	3.11	3.21	3.12	3.10	3.14
11	1.92	6.26	2.39	6.90	2.17	6.46	2.47	7.60	2.33	6.98	2.79	3.25	3.07	3.17	3.09	3.07	3.10
12	1.98	7.55	2.49	8.24	2.26	7.97	2.58	9.18	2.43	8.31	3.04	3.55	3.35	3.47	3.37	3.35	3.38
13	1.99	7.93	2.50	8.63	2.26	8.40	2.59	9.75	2.44	8.78	3.06	3.58	3.38	3.50	3.40	3.38	3.41
14	1.86	5.04	2.27	5.39	2.09	5.01	2.34	5.67	2.23	5.33	2.58	2.92	2.79	2.86	2.80	2.79	2.82
15	1.76	3.98	2.10	3.46	1.97	3.84	2.15	3.51	2.08	3.48	2.32	2.45	2.42	2.45	2.42	2.42	2.44
16	4.02	3.70	3.13	3.60	3.09	3.51	2.80	3.60	3.06	3.50	3.05	2.88	3.15	2.88	3.09	3.18	2.97
17	3.71	3.58	3.09	3.57	2.93	3.36	2.78	3.54	3.02	3.45	2.89	2.83	3.10	2.84	3.04	3.13	2.93
18	2.19	2.75	4.06	6.27	2.17	2.70	3.59	6.19	3.31	5.41	2.25	3.96	3.89	3.54	3.91	3.85	3.62
19	4.62	4.71	3.46	4.54	3.49	4.54	3.11	4.50	3.36	4.39	3.57	3.27	3.62	3.29	3.57	3.69	3.38
20	3.51	4.60	3.44	5.18	2.89	4.29	3.09	5.14	3.35	4.99	3.04	3.38	3.73	3.38	3.67	3.77	3.49
21	2.43	3.78	3.88	9.69	2.32	3.69	3.45	9.80	3.55	8.54	2.50	4.31	4.61	4.13	4.55	4.34	4.25

22	2.29	2.68	4.05	6.00	2.19	2.54	3.57	5.64	3.58	5.15	2.21	3.84	4.14	3.61	4.11	3.94	3.74
23	2.31	3.04	4.39	7.69	2.27	3.03	3.88	7.63	3.50	6.41	2.40	4.48	4.34	3.92	4.38	4.06	4.02
24	2.33	2.82	4.34	7.51	2.24	2.72	3.82	7.25	3.76	6.25	2.31	4.40	4.60	4.09	4.61	4.32	4.20
25	2.22	2.76	4.18	6.71	2.19	2.69	3.71	6.55	3.48	5.59	2.27	4.13	4.17	3.72	4.18	3.91	3.83
26	2.30	2.89	4.26	7.57	2.23	2.80	3.78	7.33	3.66	6.29	2.31	4.35	4.50	4.01	4.50	4.20	4.13
27	2.32	2.69	4.67	7.63	2.22	2.56	4.06	7.27	3.97	6.17	2.23	4.48	4.76	4.14	4.80	4.47	4.29
28	2.12	2.39	3.85	5.17	2.07	2.35	3.42	5.02	3.31	4.46	2.11	3.55	3.67	3.29	3.68	3.48	3.40
29	2.54	3.98	2.55	3.96	2.55	4.03	2.57	4.07	2.56	3.95	2.96	2.95	2.95	2.98	2.95	2.94	2.90
30	2.87	8.70	2.87	8.51	2.87	9.74	2.90	9.79	2.87	8.81	3.87	3.86	3.86	3.93	3.88	3.86	3.78
31	2.46	4.16	2.45	4.13	2.46	4.58	2.48	4.52	2.46	4.27	2.89	2.88	2.88	2.93	2.90	2.89	2.85
32	2.68	5.46	2.67	5.42	2.68	5.99	2.71	5.93	2.68	5.55	3.35	3.33	3.33	3.40	3.35	3.34	3.28
33	2.99	10.91	2.99	10.70	2.99	12.84	3.03	12.67	2.99	11.22	4.26	4.26	4.24	4.34	4.28	4.25	4.15
34	2.72	6.91	2.73	6.91	2.73	8.12	2.76	8.09	2.73	7.29	3.57	3.59	3.56	3.65	3.59	3.58	3.51
35	2.38	3.53	2.37	3.51	2.38	3.77	2.39	3.73	2.38	3.58	2.71	2.70	2.69	2.73	2.70	2.70	2.67
36	2.48	4.30	2.47	4.27	2.48	4.85	2.50	4.76	2.48	4.46	2.95	2.95	2.93	3.00	2.96	2.95	2.91
37	2.78	7.60	2.79	7.18	2.80	8.18	2.82	8.77	2.80	7.58	3.66	3.70	3.65	3.71	3.70	3.73	3.64
38	2.28	2.95	2.29	2.90	2.29	2.97	2.30	3.01	2.29	2.92	2.48	2.48	2.48	2.49	2.49	2.52	2.47

Numbers in rows are the operating conditions of the central composite design (Table 2.2). Numbering (in column 1) of compounds is presented in Section 4.2.1.

imported into MODDE [24], and DoE models were constructed using Monte Carlo simulations [25] with standard settings from the MODDE software. This step was followed by investigation of the knowledge space, the definition of the design space, robustness testing and finally the experimental realization of the predicted results, all of which are discussed in detail in the following sections. The overall combined QSRR-DoE approach is illustrated in Figure 4.4.

4.3 Results and discussion

4.3.1 Generation of DoE models

The base DoE equation was derived, which contains the three variables (pH, percentage of acetonitrile, and salt content) as well as their interaction and quadratic terms. A multiple linear regression was performed, removing terms and optimising the model based on f-tests of statistical significance for the model and p-tests for the individual coefficients. The result of this regression is shown in eq 4.1.

$$t_R = \beta_0 + \beta_1 \times X_1 + \beta_2 \times X_2 + \beta_3 \times X_3 + \beta_4 \times X_1^2 + \beta_5 \times X_2^2 + \beta_6 \times X_1 X_2 + \beta_7 \times X_1 X_3 + \beta_8 \times X_2 X_3 \quad (4.1)$$

where X_1 and X_3 are acetonitrile content and salt concentration in the mobile phase, respectively, and X_2 is pH of the aqueous phase. The values of model regression coefficients marked as $\beta_{0\text{-}8}$ with their statistical evaluations are given in Table 4.2.

A plot of experimentally measured values for retention times of the training set analytes, compared to those predicted by the DoE equations, is depicted in Figure 4.5 and demonstrates a high correlation between the predicted and experimentally observed retention times, with $R^2 \geq 0.95$. The model also passes the *f*-test at a 95% confidence level.

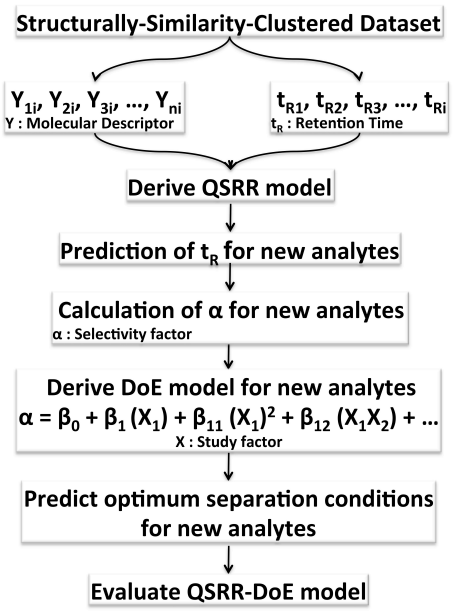

Figure 4.4. Scheme of the combined QSRR-DoE modelling approach followed in this work. The QSRR-DoE model was used to predict t_R of new analytes under new chromatographic conditions.

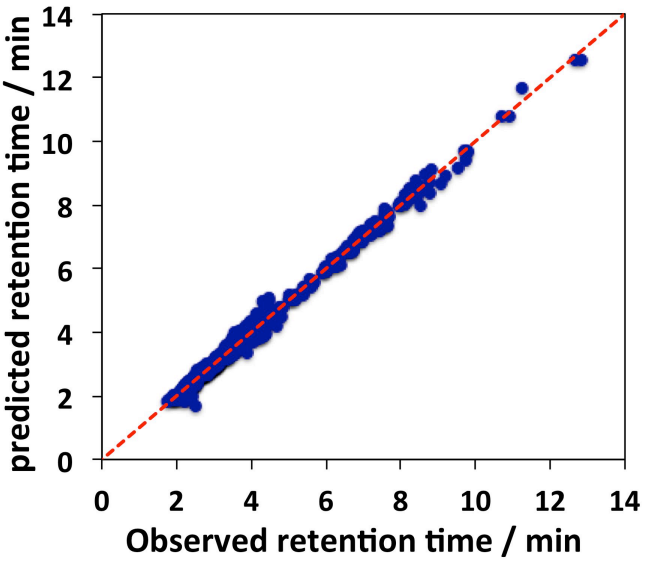

Figure 4.5. Predicted retention times *versus* measured retention times of DoE models for all compounds in the training set. A total of 646 data points is included.

Table 4.2. Coefficients of the obtained DoE models for the training set and their statistical evaluation.

nr.	β_0(p)	β_1(p)	β_2(p)	β_3(p)	β_4(p)	β_5(p)	β_6(p)	β_7(p)	β_8(p)	Q^2	R^2	R^2 adj.
1	2.997 (0.000)	2.124 (0.000)	0.280 (0.000)	0.168 (0.002)	1.383 (0.000)		0.103 (0.048)			0.989	0.997	0.995
2	3.150 (0.000)	2.525 (0.000)	0.473 (0.000)	0.116 (0.048)	1.672 (0.000)		0.323 (0.000)			0.984	0.996	0.995
3	3.006 (0.000)	2.358 (0.000)	0.169 (0.000)	0.107 (0.006)	1.573 (0.000)					0.995	0.998	0.998
4	2.970 (0.000)	1.996 (0.000)	0.211 (0.000)	0.095 (0.002)	1.240 (0.000)					0.995	0.999	0.998
5	3.128 (0.000)	2.342 (0.000)	0.261 (0.000)	0.130 (0.003)	1.486 (0.000)		0.087 (0.039)			0.993	0.998	0.997
6	3.090 (0.000)	2.295 (0.000)	0.240 (0.000)	0.124 (0.002)	1.460 (0.000)					0.994	0.998	0.997
7	3.400 (0.000)	3.037 (0.000)	0.342 (0.000)	0.204 (0.013)	2.030 (0.000)					0.983	0.995	0.993
8	2.656 (0.000)	1.027 (0.000)	0.193 (0.000)		0.491 (0.000)					0.99	0.996	0.994
9	3.330 (0.000)	2.930 (0.000)	0.291 (0.000)	0.183 (0.005)	1.937 (0.000)					0.991	0.997	0.995
10	3.120 (0.000)	2.168 (0.000)	0.300 (0.000)	0.113 (0.004)	1.325 (0.000)		0.117 (0.007)			0.993	0.998	0.997
11	3.081 (0.000)	2.291 (0.000)	0.300 (0.000)	0.133 (0.008)	1.467 (0.000)		0.125 (0.019)			0.99	0.997	0.996
12	3.362 (0.000)	2.951 (0.000)	0.324 (0.000)	0.185 (0.004)	1.938 (0.000)		0.133 (0.036)			0.99	0.997	0.996
13	3.393 (0.000)	3.170 (0.000)	0.342 (0.000)	0.207 (0.006)	2.133 (0.000)		0.151 (0.047)			0.987	0.997	0.995
14	2.798 (0.000)	1.566 (0.000)	0.200 (0.000)	0.062 (0.021)	0.924 (0.000)					0.994	0.998	0.997
15	2.418 (0.000)	0.820 (0.000)			0.415 (0.000)					0.965	0.988	0.982
16	3.010 (0.000)	0.181 (0.000)	-0.137(0.002)	-0.172(0.000)	0.392 (0.000)		0.147 (0.002)	0.131 (0.004)	0.100 (0.019)	0.630	0.955	0.920
17	2.946 (0.000)	0.197 (0.000)	-0.066(0.024)	-0.162(0.000)	0.358 (0.000)		0.118 (0.002)	0.106 (0.004)	0.082 (0.014)	0.786	0.966	0.939
18	3.745 (0.000)	0.800 (0.000)	1.201 (0.000)		0.624 (0.001)	-0.630(0.001)	0.464 (0.000)			0.925	0.973	0.96

19	3.460 (0.000)	0.464 (0.000)	-0.205(0.002)	-0.202(0.002)	0.613 (0.000)		0.168 (0.010)	0.158 (0.014)		0.684	0.963	0.935
20	3.542 (0.000)	0.792 (0.000)	0.190 (0.000)	-0.166(0.000)	0.708 (0.001)	-0.252(0.001)	0.163 (0.001)	0.079 (0.028)	0.068 (0.049)	0.977	0.994	0.988
21	4.422 (0.000)	1.986 (0.000)	1.640 (0.000)		1.557 (0.000)	-1.083(0.002)	1.180 (0.000)			0.926	0.974	0.962
22	3.885 (0.000)	0.633 (0.000)	1.120 (0.000)	-0.161(0.039)	0.532 (0.001)	-0.810(0.000)	0.410 (0.000)			0.933	0.978	0.965
23	4.170 (0.000)	1.144 (0.000)	1.500 (0.000)		0.810 (0.001)	-0.706(0.002)	0.695 (0.000)			0.931	0.975	0.964
24	4.364 (0.000)	1.005 (0.000)	1.491 (0.000)		0.697 (0.002)	-0.948(0.000)	0.704 (0.000)			0.93	0.974	0.962
25	3.976 (0.000)	0.852 (0.000)	1.316 (0.000)		0.611 (0.002)	-0.723(0.001)	0.541 (0.000)			0.929	0.973	0.961
26	4.275 (0.000)	1.064 (0.000)	1.476 (0.000)		0.750 (0.001)	-0.893(0.000)	0.712 (0.000)			0.932	0.975	0.963
27	4.470 (0.000)	0.910 (0.000)	1.610 (0.000)		0.698 (0.005)	-1.015(0.000)	0.682 (0.000)			0.919	0.969	0.955
28	3.501 (0.000)	0.463 (0.000)	1.000 (0.000)		0.422 (0.004)	-0.633(0.000)	0.297 (0.002)			0.919	0.969	0.955
29	2.950 (0.000)	0.722 (0.000)		0.023 (0.011)	0.330 (0.000)					0.998	0.999	0.998
30	3.863 (0.000)	3.118 (0.000)		0.243 (0.000)	2.130 (0.000)			0.286 (0.000)		0.996	0.998	0.997
31	2.890 (0.000)	0.934 (0.000)		0.090 (0.000)	0.508 (0.000)			0.098 (0.000)		0.996	0.998	0.998
32	3.340 (0.000)	1.493 (0.000)		0.115 (0.000)	0.838 (0.000)			0.126		0.997	0.998	0.998

					(0.000)			
33	4.256 (0.000)	4.334 (0.000)	0.404 (0.000)	3.080 (0.000)	0.482 (0.000)	0.995	0.997	0.996
34	3.580 (0.000)	2.367 (0.000)	0.252 (0.000)	1.520 (0.000)	0.293 (0.000)	0.996	0.998	0.997
35	2.700 (0.000)	0.623 (0.000)	0.051(0.000)	0.303 (0.000)	0.054 (0.000)	0.996	0.998	0.998
36	2.950 (0.000)	1.024 (0.000)	0.114 (0.000)	0.554 (0.000)	0.126 (0.000)	0.995	0.998	0.997
37	3.684 (0.000)	2.533 (0.000)	0.228 (0.003)	1.646 (0.000)	0.265 (0.003)	0.985	0.994	0.992
38	2.487 (0.000)	0.330 (0.000)	0.017 (0.024)	0.133 (0.000)		0.99	0.996	0.994

Numbering (in column) of compounds is available in Section 4.2.1. DoE model is $t_R = \beta_0 + \beta_1 \times$ acetonitrile content $+ \beta_2 \times$ pH $+ \beta_3 \times$ salt concentration $+ \beta_4 \times$ (acetonitrile content)$^2 + \beta_5 \times$ (pH)$^2 + \beta_6 \times$ (acetonitrile content $\times$ pH) $+ \beta_7 \times$ (acetonitrile content $\times$ salt concentration) $+ \beta_8 \times$ (acetonitrile content $\times$ pH). p is the significance of the variables in the model.

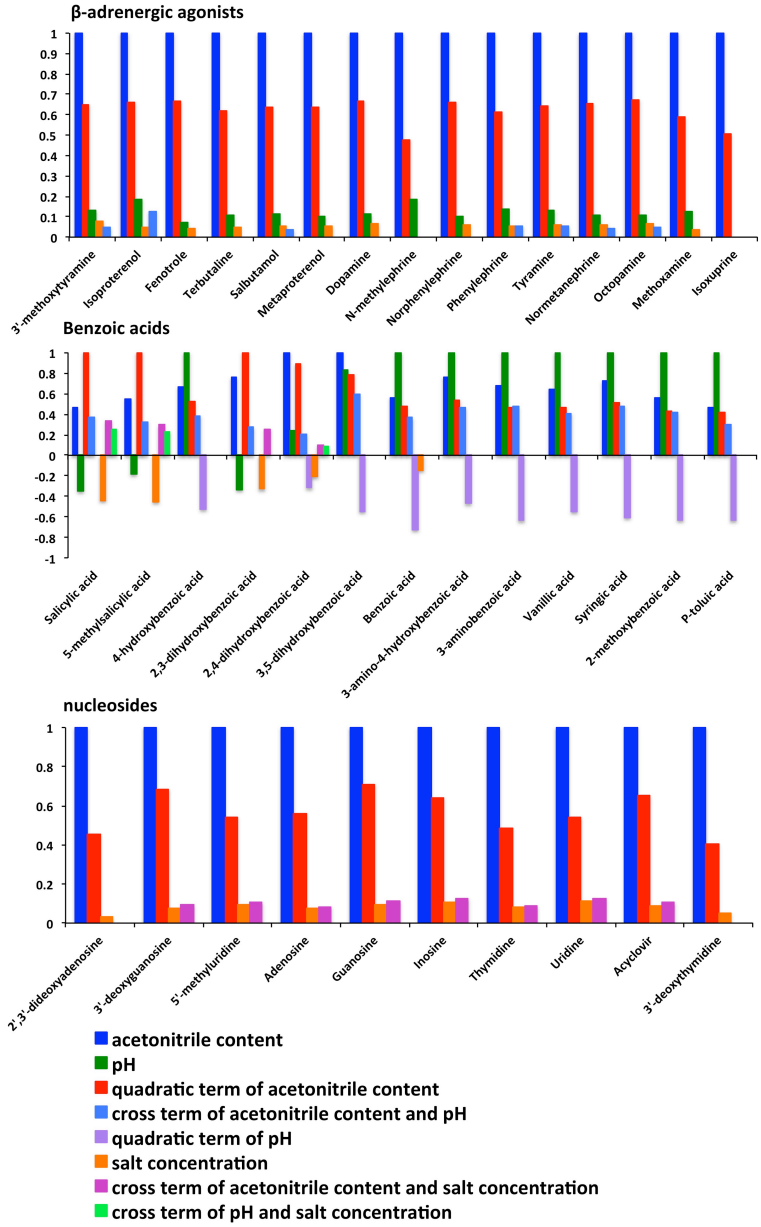

Figure 4.6. DoE model term ranking chart for each compound in the training set. In model term ranking, the variable that has the highest coefficient value is assigned a ranking of 1, and other variables are ranked based on the experimental data.

The model term ranking chart, presented in Figure 4.6 allows determination of the most significant factors in determining retention of analytes. The data in Figure 4.6 indicate that the linear and the quadratic terms of the acetonitrile content were significant factors in determining retention of the studied analytes, in agreement with the understanding that both hydrophilic partitioning as well as adsorption take part in the overall retention mechanism in HILIC mode. No ion-exchange interactions are present for the nucleosides and ion-exchange interactions are only present to a small extent for the beta-agonists. The retention of the benzoic acids on the other hand is significantly influenced by pH and ion-exchange interactions. For benzoic acids with $pK_a > 4$, a large positive coefficient was observed for the pH suggesting increased hydrophilicity of these acids due to solute deprotonation in higher pH [26, 27]. In contrast, the pH term in the DoE models obtained for benzoic acids with $pK_a < 3$ has a negative coefficient, possibly due to the effect of pH on diminishing the repulsion of these analytes from charged silanol groups on the stationary phase [26].

To evaluate experimentally the predictive power of the DoE models, retention times of the training set compounds were measured using the HILIC condition of 10 mmol L^{-1} ammonium formate (pH 4) containing 85% v/v acetonitrile, which was different from conditions used to derive the DoE models. The results include 38 measured retention times, which were compared with the predicted values and an average RMSEP% value of 9.26 was obtained as shown in Figure 4.7.

However, the original retention data used to derive the DoE models were acquired on the column when new and the observed differences between predicted and measured retention times could possibly be attributed to changes in the column behaviour due to its numerous uses (more than 2000 runs). Therefore, there is a clear need to update the retention database to

reflect changes in the retention times resulting from changes in the column over time. For this purpose, a porting procedure based on our previously developed porting methodologies [28, 29] was applied by employing three analytes (least retained, middle retained and most retained compounds) as markers to derive porting equations for each structurally-similarity based cluster. Retention data of these markers were measured experimentally on the same column used in Figure 4.7 and the ratio of new to original values was calculated and used to recalibrate the predicted retention values for all analytes in the dataset under the test conditions. This led to a substantial improvement in the match between predicted and measured results and highlights the need to use ported retention datasets. The comparison between the ported predicted retention times and those observed experimentally for the test condition is depicted in Figure 4.8 with the average RMSEP% value of 3.95. The derived porting equation for each cluster in the dataset is listed in Table 4.3.

However, while the above results show that DoE models are successful in predicting retention for new mobile phase conditions, they are unable to predict the retention times of new analytes not included in the modelling process. To overcome this drawback, we applied a two-stage approach, in which retention times of new analytes for each operating condition of the central composite design were predicted by applying the QSRR methodology described earlier, before their further use in the calculation of coefficients for the DoE models. In this way, the DoE models for the new analytes could be used to predict retention times for any mobile phase composition in the design space.

4.3.2 Combined QSRR and DoE modelling

QSRR models were generated, as described in the Model Generation section, using the experimental chromatographic retention times, with DFT-computed molecular descriptors calculated for three datasets classified according to structural similarity: 15 β-adrenergic agonists, 13 benzoic acids and 10 nucleosides. The predictivities of the GA-PLS models obtained for all 17 chromatographic conditions are presented in Table 4.4 with internal validation by RMSECV (equation 2.5) giving an average value of 6.08% over all analysed HILIC conditions, which indicates that the predicted values for the retention of all tested analytes were in relatively good agreement with the experimental data.

The reliability of the QSRR models was tested by applying to the training sets Y-randomization tests with 1000 iterations and randomly assigned retention times. The result confirmed the reliability of the models, given that RMSECV values for the actual dataset were shown to be significantly lower than those obtained for the same dataset with randomized retention data (Figures 4.9 – 4.11).

External validation data sets were used to illustrate the utility of our approach to predict the retention time of an unstudied set of compounds for a set of experiments covering a large body of chromatographic conditions. External validation of the predictive power of the QSRR models was evaluated using 4 test analytes in each cluster, which had not been utilised in either molecular descriptor feature selection or model generation. These test analytes comprised β-adrenergic agonists (adrenaline, noradrenaline, ritudrine, synephrine), benzoic acids (3-hydroxybenzoic acid, 2,5-dihydroxybenzoic acid, 4-aminobenzoic acid, 4-aminosalicylic acid) and nucleosides (2'-deoxyadenosine, 2'-deoxyguanosine, 2'-deoxyuridine, 2'-deoxyinosine).

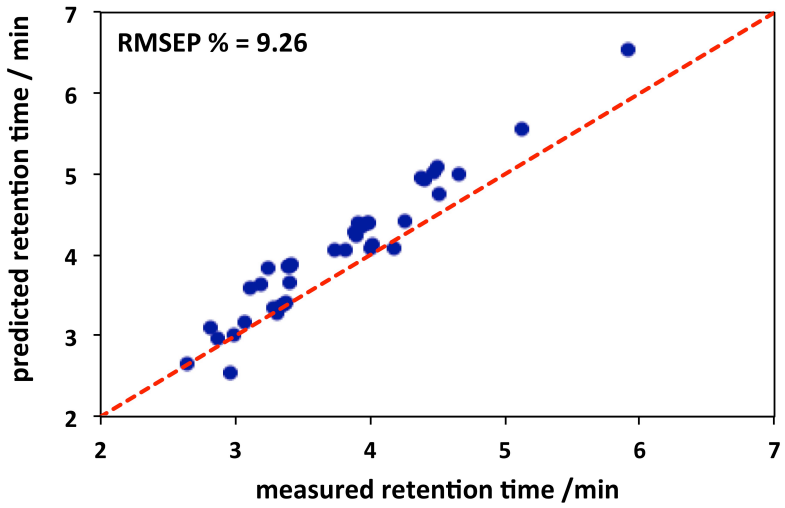

Figure 4.7. Experimental values versus predicted values of external validation of DoE models for analysed compounds under a never analysed chromatographic condition: 10 mmol L^{-1} ammonium formate (pH 4) containing 85% v/v acetonitrile.

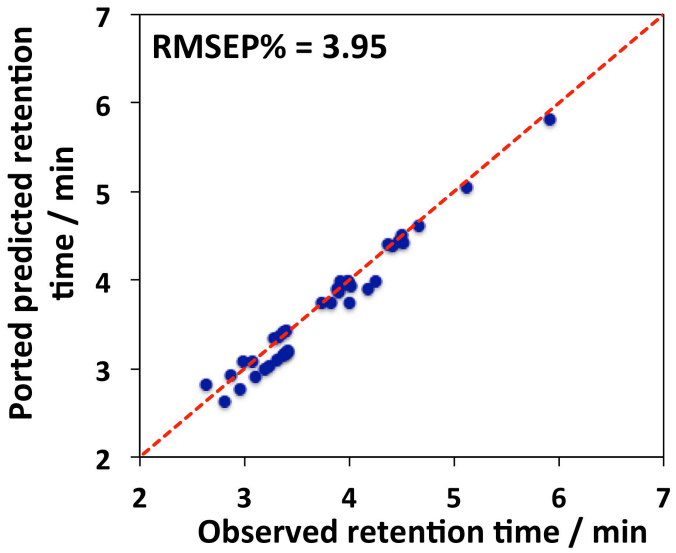

Figure 4.8. Experimental retention times *versus* ported predicted retention times for external validation of DoE models under a never analysed chromatographic condition: 10 mmol L^{-1} ammonium formate (pH 4) containing 85% v/v acetonitrile.

Table 4.3. Summary of porting equations on the structurally-similarity based classified datasets.

Dataset	Porting equation
β-agonists	$t_{Rported} = 0.716\ t_{Rmeasured}^{a} + 0.872$
Benzoic acids	$t_{Rported} = 0.931\ t_{Rmeasured} + 0.015$
nucleosides	$t_{Rported} = 0.671\ t_{Rmeasured} + 1.100$

a is measured retention time when the column was new.

Table 4.4. Predictive performance of GA-PLS models on internal validation.

	RMSECV (%) datasets:		
nr.	β-agonists	benzoic acids	nucleosides
1	0.43	5.83	0.64
2	3.50	8.61	4.40
3	0.64	12.01	0.63
4	5.71	38.05	6.01
5	0.59	3.79	0.87
6	4.74	6.59	6.46
7	0.52	5.70	0.83
8	3.93	34.91	4.88
9	0.64	9.88	0.66
10	3.51	22.13	4.26
11	1.22	3.56	0.96
12	1.07	11.47	1.78
13	2.81	19.55	2.12
14	1.39	8.84	1.49
15	1.44	16.58	1.24
16	1.02	12.84	1.63
17	1.09	15.35	1.42

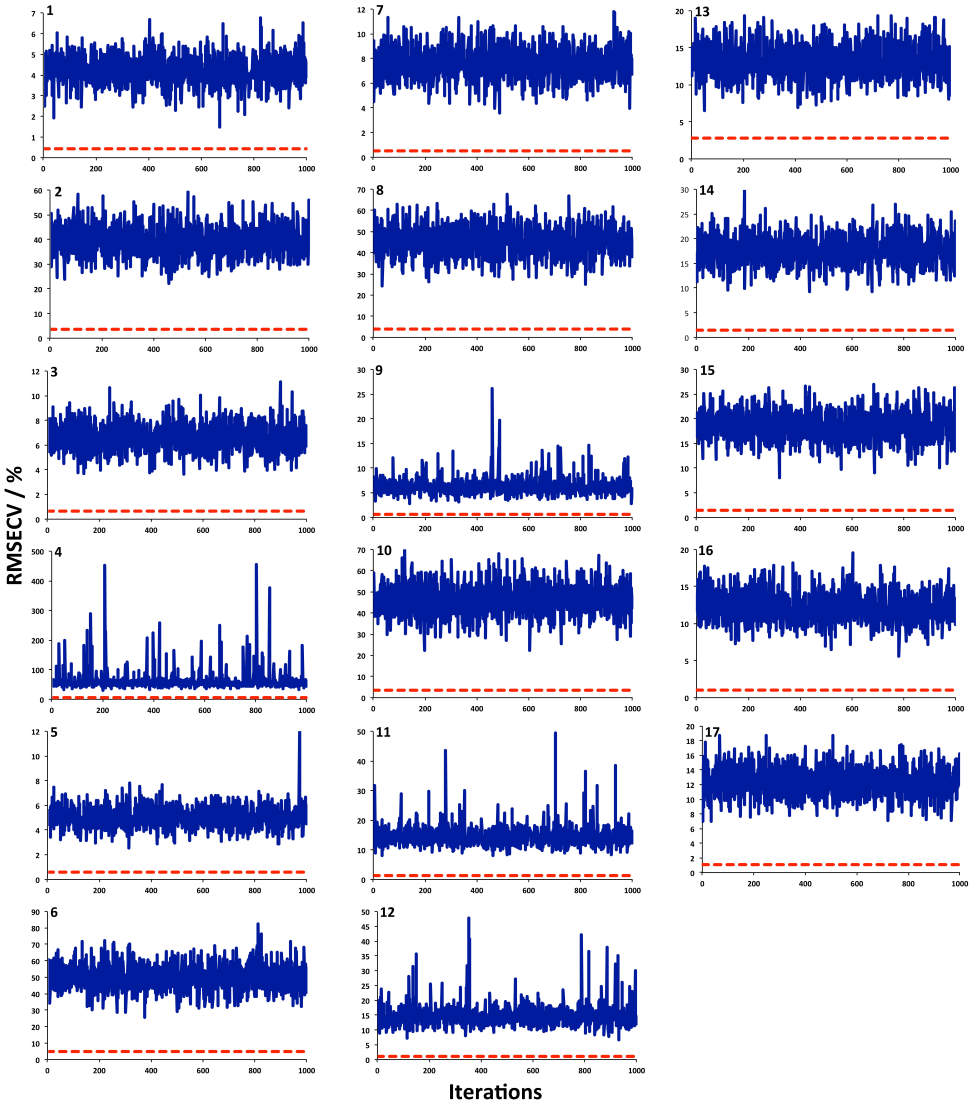

Figure 4.9. Y-randomization plots of β-adrenergic agonists dataset on 17 operating chromatographic conditions of the studied central composite design. Y-randomized data (royal blue lines), and the actual data (red line).

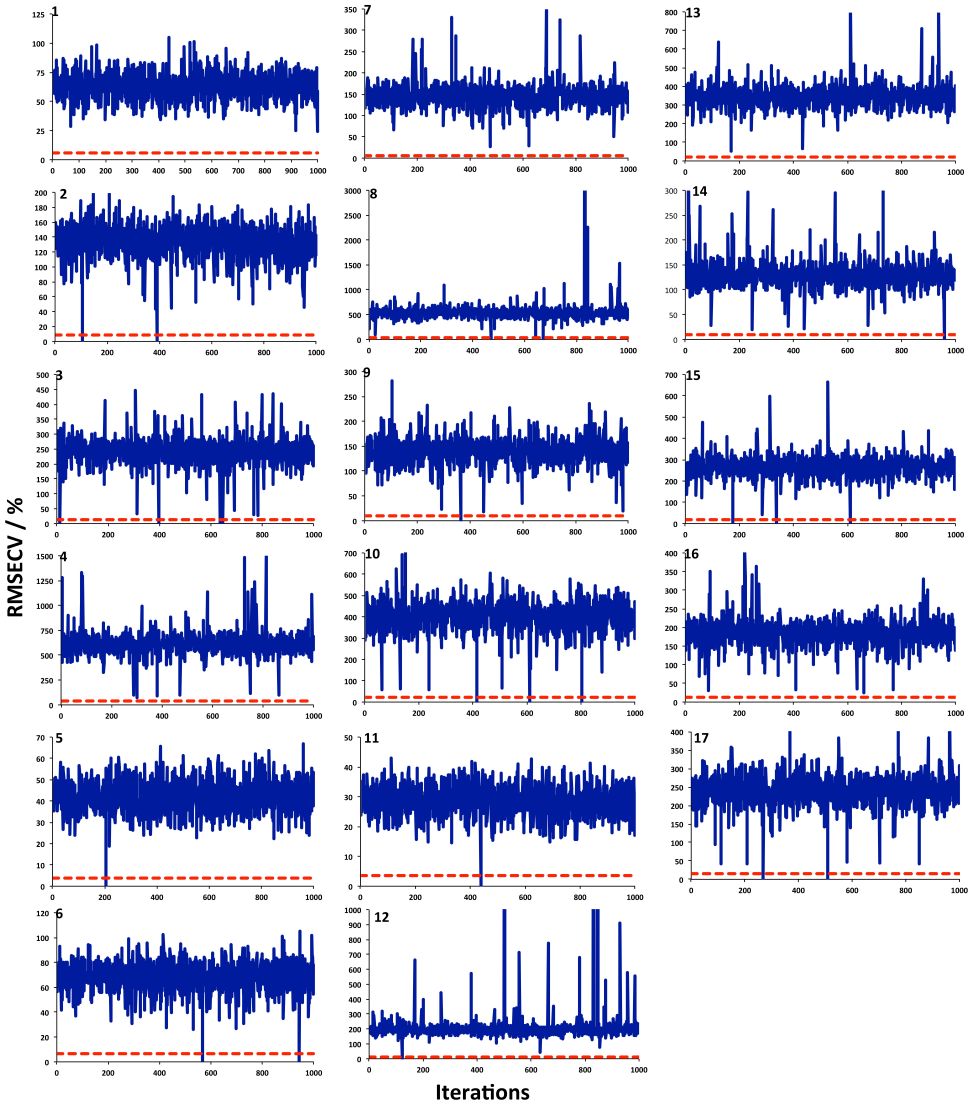

Figure 4.10. Y-randomization plots of benzoic acids dataset on 17 operating chromatographic conditions of the studied central composite design. Y-randomized data (royal blue lines), and the actual data (red line).

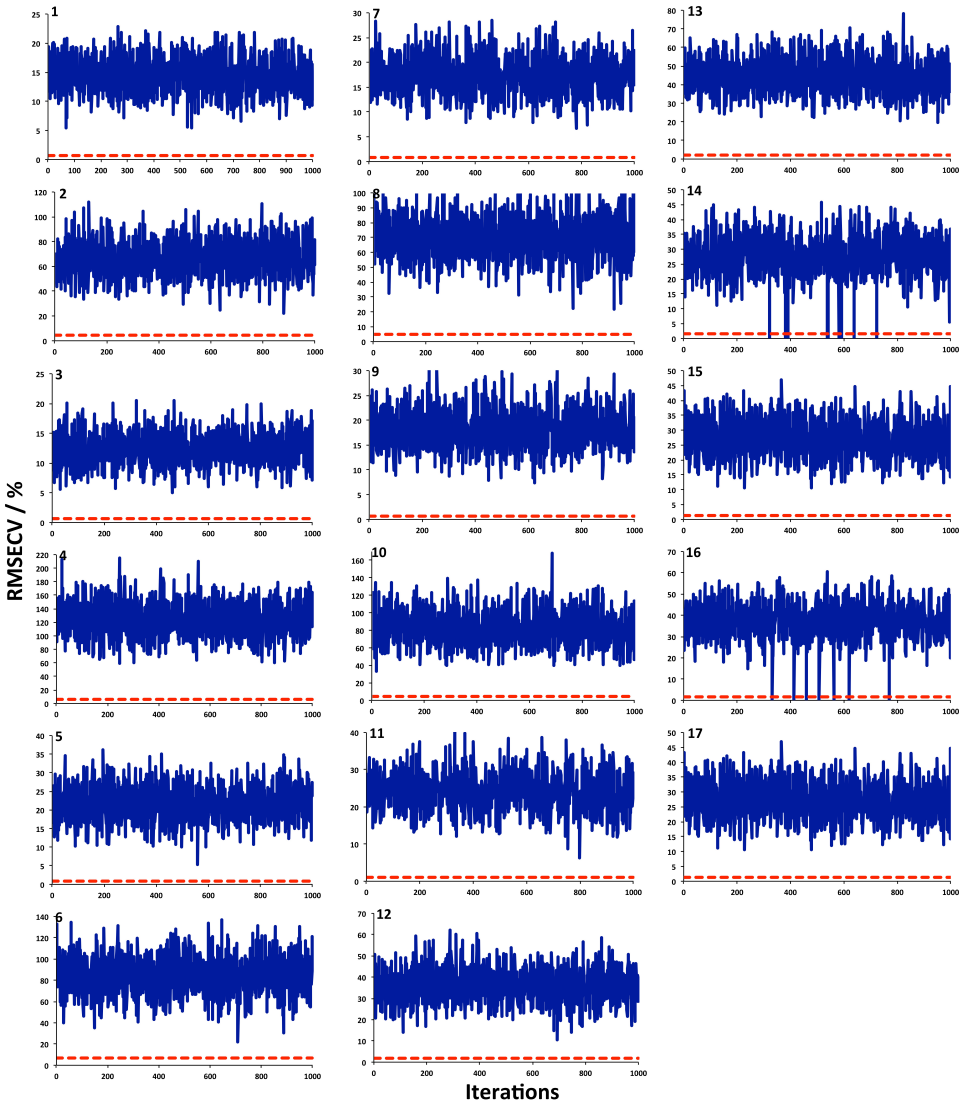

Figure 4.11. Y-randomization plots of nucleosides dataset on 17 operating chromatographic conditions of the studied central composite design. Y-randomized data (royal blue lines), and the actual data (red line).

The appropriate descriptor values for each analyte in the test sets were inserted into the GA-PLS model equation, and the respective retention times were calculated and compared to measured experimental retention times. A comparison between the predicted retention times and those observed experimentally for each test compound using the GA-PLS model is depicted in Figure 4.12, while figures of merit are summarised in Table 4.5. As seen in Table 4.5, in general, the GA-PLS models generated for the test compounds over all chromatographic conditions used for data acquisition in the design space were very well correlated with the experimental data, presenting an average RMSEP% (equation 2.4) value for all analytes over all experimental conditions of 6.74%. The prediction errors for the benzoic acids in a few conditions (Table 4.5) are higher than desired, possibly due to a complex, multi-modal retention mechanism for this cluster. The ability to predict retention of new compounds for the mobile phase compositions used to define the design space opened the possibility of applying DoE modelling and hence prediction of retention of these new compounds under new mobile phase conditions. The above GA-PLS step was followed by the calculation of coefficient values for the DoE models from the corresponding predicted retention times. Thus, from the QSRR data, we obtained DoE models for each analyte in the external validation set. Coefficients of the obtained DoE models and their statistical evaluations for targets are presented in Table 4.6 and the experimental and predicted retention times of targets are presented in Table 4.7. The performance of the obtained QSRR-DoE models was evaluated by prediction of the retention times of the test analytes under a chromatographic condition of 10 mmol L^{-1} ammonium formate (pH 4) containing 85% v/v acetonitrile (which had not been used in any modelling procedure). A plot of experimental data and model

Table 4.5. QSRR models performance summary

		RMSEP / % in datasets:	
nr.	β-agonists	benzoic acids	nucleosides
1	1.41	5.35	1.38
2	6.43	4.30	11.43
3	1.65	2.36	1.40
4	9.73	13.41	8.70
5	1.63	2.89	0.81
6	7.21	5.36	18.21
7	1.55	7.40	1.42
8	6.76	16.22	8.35
9	1.70	13.69	0.90
10	4.70	30.53	11.57
11	2.90	4.76	3.55
12	2.83	5.04	4.72
13	3.44	11.10	2.87
14	3.18	23.92	2.74
15	3.00	10.98	5.75
16	3.30	12.47	2.34
17	2.87	20.64	2.77

Method abbreviation explained in the text. nr. is the number assigned to each experimental condition of the applied central composite design.

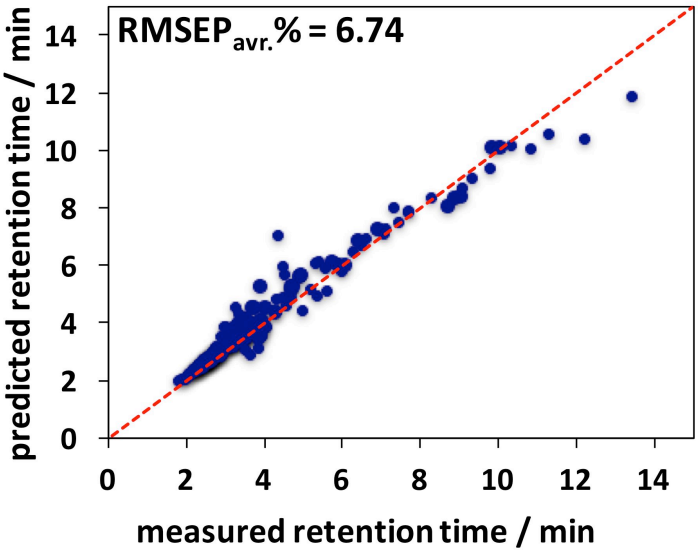

Figure 4.12. Predictive ability of GA-PLS models for external validation sets of β-adrenergic agonists, benzoic acids and nucleosides over all 17 experimental conditions of the central composite design used in this study. RMSEP$_{avr.}$% is the average value of RMSEP% of test analytes over all studied conditions. A total of 204 data points is included.

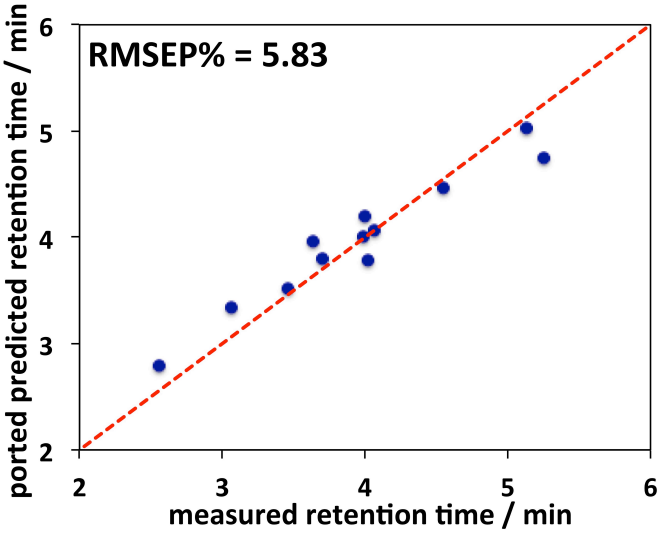

Figure 4.13. Experimental retention times *versus* ported predicted retention times for combined QSRR-DoE models for never analysed compounds under a never analysed condition: 10 mmol L^{-1} ammonium formate (pH 4) containing 85% v/v acetonitrile

predictions for the test analytes is shown in Figure 4.13 and confirms good QSRR-DoE model performance in predicting the retention times of the analytes.

4.3.3 Prediction of the optimal separation conditions by applying QSRR-DoE-QbD methodology

In the previous sections, the prediction abilities of the obtained QSRR-DoE models were investigated. The obtained result shows clearly that the combined QSRR-DoE methodology reproduces the experimentally obtained retention times with acceptable RMSEP values. This is one of the central results of this work: the combined QSRR-DoE method is capable of using the chemical structures of new analytes to predict retention times under a new HILIC condition with an acceptably low error. These results permit the combination of QSRR-DoE and QbD to determine the robustness of the separation over a chosen design space of mobile phase compositions. To investigate this approach, the optimal separation conditions for 4 test analytes from each analyte cluster (see above for analyte identities) were predicted computationally, based on the chemical structures of the analytes and employing QSRR-DoE and QbD.

Retention data predicted from the QSRR-DoE modelling were converted to selectivity values (α) for all analyte pairs and imported into MODDE 10 software [24] and the investigated experimental domain was explored to identify the optimal separation conditions. In this case the performance was expressed as the percentage risk of failure of the separation, with failure being defined as $\alpha<1.15$ for any analyte pair. In accordance with QbD principles, the defined design space should not only provide the required optimal level of performance in terms of separation of test analytes, but also identify the robust regions to gain assurance in quality of the developed

Table 4.6. Coefficients of the obtained DoE models for the test sets and their statistical evaluation.

analytes	β_0(p)	β_1(p)	β_2(p)	β_3(p)	β_4(p)	β_5(p)	β_6(p)	β_7(p)	Q^2	R^2	R^2 adj.
Adrenaline	3.418 (0.000)	3.190 (0.000)	0.447 (0.000)	0.209 (0.008)	2.148 (0.000)		0.268 (0.004)		0.986	0.996	0.995
Noradrenaline	3.674 (0.000)	3.960 (0.000)	0.436 (0.000)	0.259 (0.013)	2.748 (0.000)		0.238 (0.034)		0.984	0.996	0.994
Ritudrine	2.934 (0.000)	1.762 (0.000)			1.034 (0.000)		-0.245 (0.001)		0.975	0.993	0.989
Synephrine	3.200(0.000)	2.353 (0.000)	0.319 (0.000)	0.133 (0.019)	1.463 (0.000)		0.132 (0.033)		0.987	0.996	0.994
3-Hydroxybenzoic acid	4.123 (0.000)	1.270 (0.000)	1.346 (0.000)		1.139 (0.002)	-1.081 (0.003)	0.813 (0.001)		0.725	0.964	0.928
2,5-Dihydroxybenzoic acid	4.048 (0.000)	1.025 (0.000)			1.276 (0.000)	-0.913 (0.003)	0.413 (0.014)		0.674	0.941	0.883
4-Aminobenzoic acid	3.617 (0.000)	0.296 (0.004)	0.830 (0.000)			-0.603 (0.002)			0.776	0.955	0.91
4-Aminosalicylic acid	3.258 (0.000)	0.530 (0.010)	0.487 (0.010)		0.608 (0.010)				0.589	0.899	0.797
2'-deoxyadenosine	3.225 (0.000)	1.350 (0.000)		0.094 (0.001)	0.761 (0.000)			0.103 (0.001)	0.995	0.997	0.996
2'-deoxyguanosine	4.995 (0.000)	3.034 (0.000)							0.655	0.805	0.76
2'-deoxyuridine	3.226 (0.000)	1.050 (0.000)							0.713	0.843	0.807
2'-deoxyinosine	4.050 (0.000)	1.863 (0.000)							0.724	0.847	0.812

DoE model is $t_R = \beta_0 + \beta_1 \times$ acetonitrile content $+ \beta_2 \times$ pH $+ \beta_3 \times$ salt concentration $+ \beta_4 \times$ (acetonitrile content)$^2 + \beta_5 \times$ (pH)$^2 + \beta_6 \times$ (acetonitrile content $\times$ pH) $+ \beta_7 \times$ (acetonitrile content $\times$ salt concentration) $+ \beta_8 \times$ (acetonitrile content $\times$ pH). p is the significance of the variables in the model.

Table 4.7. The experimental and predicted retention times of test compounds for each operating condition of the central composite design by using GA-PLS models.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Adrenaline																	
$t_{Rmeasured}$	2.00	7.75	2.56	9.38	2.28	8.33	2.65	10.39	2.47	9.12	3.09	3.69	3.45	3.58	3.48	3.45	3.46
$t_{Rpredicted}$	2.00	7.79	2.54	8.98	2.28	8.26	2.62	10.11	2.44	8.64	3.07	3.63	3.40	3.53	3.43	3.42	3.43
Noradrenaline																	
$t_{Rmeasured}$	2.05	9.81	2.63	12.26	2.35	10.86	2.73	13.45	2.54	11.32	3.32	3.96	3.70	3.85	3.75	3.71	3.71
$t_{Rpredicted}$	2.06	9.28	2.64	10.32	2.35	9.99	2.71	11.80	2.55	10.52	3.30	3.86	3.75	3.79	3.67	3.65	3.69
Ritudrine																	
$t_{Rmeasured}$	1.84	5.43	2.25	5.66	2.08	5.37	2.33	6.01	2.22	5.63	2.57	2.90	2.78	2.84	2.79	2.76	2.81
$t_{Rpredicted}$	1.89	6.06	2.32	5.07	2.14	6.01	2.40	5.72	2.28	5.81	2.72	3.04	2.95	3.01	2.94	2.93	2.95
Synephrine																	
$t_{Rmeasured}$	1.94	6.32	2.43	7.10	2.19	6.53	2.52	7.74	2.36	7.15	2.85	3.36	3.16	3.27	3.17	3.15	3.18
$t_{Rpredicted}$	1.96	6.40	2.46	7.06	2.21	6.61	2.53	7.83	2.39	7.19	2.87	3.36	3.22	3.27	3.20	3.20	3.27
3-Hydroxybenzoic acid																	
$t_{Rmeasured}$	2.34	3.09	3.99	7.50	2.25	2.97	3.52	7.36	3.58	6.67	2.34	4.12	4.37	3.85	4.33	4.35	4.03
$t_{Rpredicted}$	2.34	3.15	3.92	7.42	2.25	2.95	3.42	7.93	3.71	6.88	2.35	3.81	4.31	3.59	4.22	4.38	4.04
2,5-Dihydroxybenzoic acid																	
$t_{Rmeasured}$	4.17	4.77	3.20	4.59	3.22	4.62	2.89	4.64	3.14	4.46	3.38	3.15	3.45	3.17	3.39	3.50	3.25
$t_{Rpredicted}$	4.32	4.80	3.27	5.58	3.34	4.80	2.98	5.91	3.93	7.00	3.43	3.12	4.06	4.02	4.10	3.56	4.22
4-Aminosalicylic acid																	
$t_{Rmeasured}$	2.80	3.59	3.66	5.40	2.50	3.32	3.27	5.24	3.52	5.01	2.58	3.58	3.96	3.55	3.89	3.91	3.68

$t_{Rpredicted}$	3.10	3.59	3.63	4.84	2.59	3.55	3.49	5.11	3.34	4.36	2.79	3.42	4.05	3.00	3.35	3.08	2.84
4-Aminobenzoic acid																	
$t_{Rmeasured}$	2.14	2.47	3.71	4.34	2.12	2.48	3.36	4.27	2.92	3.78	2.20	3.39	3.23	3.01	3.25	3.07	3.06
$t_{Rpredicted}$	2.15	2.67	3.80	4.74	2.15	2.62	3.79	4.38	3.47	3.92	2.27	3.45	3.82	3.81	3.52	3.45	3.79
2'-Deoxyadenosine																	
$t_{Rmeasured}$	2.60	4.73	2.60	4.71	2.60	4.93	2.63	4.99	2.61	4.75	3.16	3.15	3.14	3.19	3.16	3.15	3.09
$t_{Rpredicted}$	2.63	5.10	2.63	5.21	2.63	5.56	2.66	5.61	2.63	5.19	3.25	3.18	3.21	3.27	3.25	3.25	3.16
2'-Deoxyguanosine																	
$t_{Rmeasured}$	2.89	8.93	2.88	8.75	2.89	9.88	2.92	10.10	2.89	9.08	3.95	3.95	3.93	4.01	3.95	3.93	3.85
$t_{Rpredicted}$	2.91	8.26	2.82	8.01	2.86	10.03	2.87	10.02	2.85	8.33	3.76	3.60	4.01	3.85	3.52	3.51	3.71
2'-Deoxyuridine																	
$t_{Rmeasured}$	2.39	3.73	2.40	3.62	2.41	3.91	2.41	4.06	2.40	3.76	2.77	2.77	2.75	2.79	2.78	2.82	2.75
$t_{Rpredicted}$	2.45	4.48	2.42	3.97	2.44	5.23	2.45	4.49	2.42	4.49	2.87	2.86	2.88	2.85	2.83	2.88	2.84
2'-Deoxyinosine																	
$t_{Rmeasured}$	2.63	5.95	2.65	5.75	2.66	6.45	2.67	6.96	2.65	6.09	3.35	3.39	3.34	3.40	3.39	3.43	3.33
$t_{Rpredicted}$	2.66	5.94	2.65	6.06	2.67	6.80	2.66	7.20	2.65	5.93	3.43	3.37	3.40	3.34	3.34	3.37	3.39

Numbers in the first row are the operating conditions of the central composite design (Table 2.2).

methods. To assess this robustness, Monte Carlo simulations [25] were performed to propagate the model uncertainty, thus allowing the probability of the selected performance criteria at the optimal conditions and consequently the design space of the analytical method to be estimated. Areas of the design space where the percentage risk of failure was <2% were considered to provide a robust separation. Figure 4.14 shows plots of the percentage risk of failure over the tested range of acetonitrile compositions and pH for the test analytes in each cluster. For simplicity in displaying these data, a fixed salt concentration was used for each cluster and this is justified after considering that the salt concentration exerted only a minor effect on separation selectivity as calculated by the MODDE software. The optimal mobile phase composition is indicated in each plot and the separation obtained for each optimal composition is also shown, together with the retention time predicted from the modelling (red lines). In the case of β-adrenergic agonists, the mobile phase pH was chosen to be slightly lower than the national optimum to reduce peak tailing of adrenaline and noradrenaline.

These results show that coupling QbD methodology with retention time predictions generated using QSRR-DoE is a powerful tool to predict the optimal separation conditions for new compounds, based on their chemical structures.

4.3.4 Conclusions

A novel QbD methodology using a compound-classification-based QSRR modelling approach in combination with DoE principles was demonstrated using HILIC separation examples. The DoE model was used to relate analyte retention time to mobile-phase pH, acetonitrile content and salt concentration. A cluster-based QSRR model was generated to create a

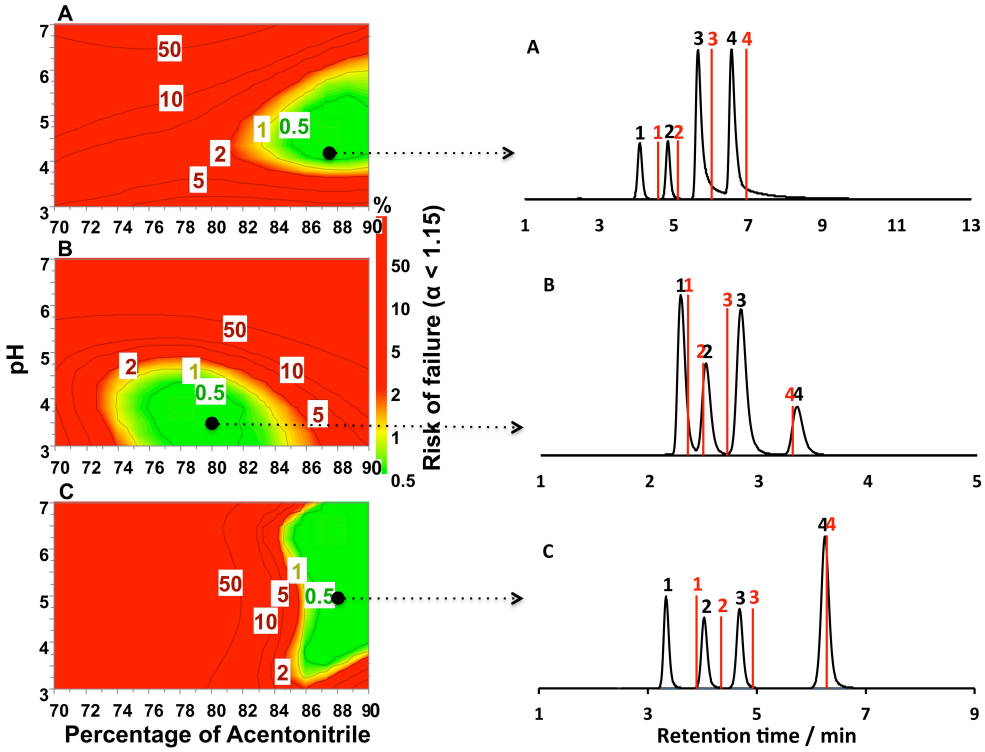

Figure 4.14. Representation of the design space of pH *versus* acetonitrile content in the mobile phase, setting the salt concentration in the mobile phase at 18.7 mmol L^{-1} for β-adrenergic agonists (A), 13.3 mmol L^{-1} for benzoic acids (B), and 16.7 mmol L^{-1} for nucleosides (C). The risk of failure map is shown for the performance criteria α with acceptance limits $\alpha \leq 1.15$. The design space is considered to be the area corresponding to a 2% risk of failure and the dots mark the mobile phase composition used to evaluate the predictive power of the models. The notional optimal mobile phase composition is used for (B) and (C), however for (A) a lower pH was used to improve peak shape. Experimental chromatograms (black peaks) corresponding to the selected working point: 18.7 mmol L-1 ammonium formate (pH 4.1) containing 87.5% v/v acetonitrile for β-adrenergic agonists (A), 13.3 mmol L-1 ammonium formate (pH 3.5) containing 80% v/v acetonitrile for benzoic acids (B), and 16.7 mmol L^{-1} ammonium formate (pH 5) containing 88% v/v acetonitrile for nucleosides (C). The red lines represent the predicted retention times from the QSRR-DoE models under the selected working condition. Numbering of test compounds in graph A: 1, ritudrine; 2, synephrine; 3, adrenaline; 4, noradrenaline; in graph B, 1, 4-aminobenzoic acid; 2, 3-hydroxybenzoic acid; 3, 4-aminosalicylic acid; 4, 2,5-dihydroxybenzoic acid; in graph C, 1, 2'-deoxyuridine; 2, 2'-deoxyadenosine; 3, 2'-deoxyinosine; 4, 2'-deoxyguanosine.

relationship between the retention data and the chemical structures of the analytes. Extensive validation of the QSRR and DoE models demonstrated excellent predictive power. A combination of these cluster-based QSRR and DoE models was used to successfully predict the retention times of new analytes under new chromatographic conditions. The calculation of the separation selectivity using the QSRR-DoE-computed retention times of targets allowed identification of the design space where the maximum peak selectivity between the test probes was located. Finally, the optimal working conditions were selected and the high level of agreement between theoretical predictions and experimental chromatography proved the adequacy of the developed method.

4.4 References

[1] T. Bolanča, Š. Ukić, M. Novak, M. Rogošić, Computer assisted method development in liquid chromatography, Croat. Chem. Acta, 87 (2014) 111-122.

[2] ICH Q8 (R2) – Guidance for Industry, Pharmaceutical Development, 2009.

[3] E. Rozet, P. Lebrun, P. Hubert, B. Debrus, B. Boulanger, Design spaces for analytical methods, TrAC, Trends Anal. Chem., 42 (2013) 157-167.

[4] L.V. Candioti, M.M. De Zan, M.S. Camara, H.C. Goicoechea, Experimental design and multiple response optimisation. Using the desirability function in analytical methods development, Talanta, 124 (2014) 123-138.

[5] I. Molnar, Computerized design of separation strategies by reversed-phase liquid chromatography: development of DryLab software, J. Chromatogr. A, 965 (2002) 175-194.

[6] M. Jovanovic, T. Rakic, A. Tumpa, B. Jancic Stojanovic, Quality by Design approach in the development of hydrophilic interaction liquid

chromatographic method for the analysis of iohexol and its impurities, J. Pharm. Biomed. Anal., 110 (2015) 42-48.
[7] S. Orlandini, B. Pasquini, M. Del Bubba, S. Pinzauti, S. Furlanetto, Quality by design in the chiral separation strategy for the determination of enantiomeric impurities: development of a capillary electrophoresis method based on dual cyclodextrin systems for the analysis of levosulpiride, J. Chromatogr. A, 1380 (2015) 177-185.
[8] T. Tol, N. Kadam, N. Raotole, A. Desai, G. Samanta, A simultaneous determination of related substances by high performance liquid chromatography in a drug product using quality by design approach, J. Chromatogr. A, 1432 (2016) 26-38.
[9] R. Kaliszan, QSRR: quantitative structure-(chromatographic) retention relationships, Chem. Rev., 107 (2007) 3212-3246.
[10] K. Muteki, J.E. Morgado, G.L. Reid, J. Wang, G. Xue, F.W. Riley, J.W. Harwood, D.T. Fortin, I.J. Miller, Quantitative structure retention relationship models in an analytical quality by design framework: simultaneously accounting for compound properties, mobile-phase conditions, and stationary-phase properties, Ind. Eng. Chem. Res., 52 (2013) 12269-12284.
[11] P. Wiczling, R. Kaliszan, How Much Can We Learn from a Single Chromatographic Experiment? A Bayesian Perspective, Anal. Chem., 88 (2016) 997-1002.
[12] P. Wiczling, L. Kubik, R. Kaliszan, Maximum A Posteriori Bayesian Estimation of Chromatographic Parameters by Limited Number of Experiments, Anal. Chem., 87 (2015) 7241-7249.
[13] T. Bączek, K. Macur, L. Bober, Rapid HPLC Method Development of Polynuclear Aromatic Hydrocarbons Separation Based on Quantitative Structure Retention Relationships, J. Liq. Chromatogr. Relat. Technol., 32 (2009) 668-679.

[14] C. Wang, M.J. Skibic, R.E. Higgs, I.A. Watson, H. Bui, J. Wang, J.M. Cintron, Evaluating the performances of quantitative structure-retention relationship models with different sets of molecular descriptors and databases for high-performance liquid chromatography predictions, J. Chromatogr. A, 1216 (2009) 5030-5038.

[15] E. Tyteca, M. Talebi, R. Amos, S.H. Park, M. Taraji, Y. Wen, R. Szucs, C.A. Pohl, J.W. Dolan, P.R. Haddad, J. Chromatogr. A, 1486 (2016) 50-58.

[16] P. Willett, Similarity methods in chemoinformatics, Annu. Rev. Inf. Sci. Technol., 43 (2009) 1-117.

[17] B. Dejaegher, Y. Vander Heyden, HILIC methods in pharmaceutical analysis, J. Sep. Sci., 33 (2010) 698-715.

[18] D. Butina, Unsupervised Data Base Clustering Based on Daylight's Fingerprint and Tanimoto Similarity: A Fast and Automated Way To Cluster Small and Large Data Sets, J. Chem. Inf. Model., 39 (1999) 747-750.

[19] in, For details on the hashed ChemAxon fingerprint developed by ChemAxon see https://docs.chemaxon.com/display/CD/Chemical+Hashed+Fingerprint [accessed January 2016].

[20] Holliday, Grouping of coefficients for the calculation of inter-molecular similarity and aissimilarity using 2D fragment bit-strings, Comb. Chem. High Throughput Screening, 5 (2002).

[21] R.E. Bruns, I.S. Scarminio, B.B. Neto, Statistical Design—Chemometrics, Elsevier, Amsterdam, 2006.

[22] L. Wang, H. Qu, Development and optimisation of SPE-HPLC-UV/ELSD for simultaneous determination of nine bioactive components in Shenqi Fuzheng Injection based on Quality by Design principles, Anal. Bioanal. Chem., 408 (2016) 2133-2145.

[23] D. Felix-Urquidez, M. Perez-Urquiza, J.B. Valdez Torres, J. Leon-Felix, R. Garcia-Estrada, A. Acatzi-Silva, Development, optimisation, and evaluation of a duplex droplet digital PCR assay To quantify the T-nos/hmg copy number ratio in genetically modified maize, Anal. Chem., 88 (2016) 812-819.

[24] MODDE v. 10. MKS Umetrics AB, in, Sweden.

[25] M.Á. Herrador, A.G. Asuero, A.G. González, Estimation of the uncertainty of indirect measurements from the propagation of distributions by using the Monte-Carlo method: An overview, Chemom. Intell. Lab. Syst., 79 (2005) 115-122.

[26] G. Greco, S. Grosse, T. Letzel, Study of the retention behaviour in zwitterionic hydrophilic interaction chromatography of isomeric hydroxy- and aminobenzoic acids, J. Chromatogr. A, 1235 (2012) 60-67.

[27] Y. Guo, S. Gaiki, Retention and selectivity of stationary phases for hydrophilic interaction chromatography, J. Chromatogr. A, 1218 (2011) 5920-5938.

[28] S.H. Park, R.A. Shellie, G.W. Dicinoski, G. Schuster, M. Talebi, P.R. Haddad, R. Szucs, J.W. Dolan, C.A. Pohl, Enhanced methodology for porting ion chromatography retention data, J. Chromatogr. A, 1436 (2016) 59-63.

[29] B.K. Ng, R.A. Shellie, G.W. Dicinoski, C. Bloomfield, Y. Liu, C.A. Pohl, P.R. Haddad, Methodology for porting retention prediction data from old to new columns and from conventional-scale to miniaturised ion chromatography systems, J. Chromatogr. A, 1218 (2011) 5512-5519.

5 Use of dual-filtering to create training sets leading to improved accuracy in quantitative structure-retention relationship modelling for hydrophilic interaction liquid chromatographic systems

5.1 Introduction

The motivation behind the use of quality by design (QbD) principles in conjunction with HPLC method development has been the desire to develop more robust and reliable analytical methods with minimal time and resource effort [1-5]. To this end, theoretical approaches have been employed for method development to propose different retention models that relate retention time/factor of an analyte to properties of the stationary phase, the eluent, and the analyte itself [6-8]. An important computational approach in predicting retention times in HPLC method development is the quantitative structure-retention relationship (QSRR) method, which correlates the retention time of an analyte to its chemical structure [9].

Although QSRR modelling methods have been used for more than 40 years, the retention time prediction accuracy of such models is not yet sufficient to support HPLC method development [9]. As a consequence, QSRR methodology is still an active research area. A key component that can potentially lead to improvement in QSRR modelling accuracy is the incorporation of appropriate molecular descriptors into the models [5]. A large number of different kinds of molecular descriptors have been reported in QSRR studies, e.g. physicochemical properties [10-12], solvatochromic descriptors [13, 14], quantum-chemical indices [15], 2D autocorrelation indices [16], GEometry, Topology, Atom-Weights AssemblY (GETAWAY) descriptors [17] and gonane topological weighted fingerprints [18]. In addition, several feature selection approaches to capture the most

informative molecular descriptors with the goal of producing more predictive models have been reported [6, 19, 20]. Comparison studies demonstrate that the well-known genetic algorithm (GA) method performs better than other possible feature selection mechanisms [20, 21].

Another strategy to enhance the level of confidence in QSRR methodology is the use of the concept of molecular similarity in compound-classification *prior to* QSRR modelling. The essence of the similarity concept is that structurally similar compounds are more likely to exhibit similar properties [22]. Classification according to similarity has proven to be a powerful tool in quantitative structure-property (activity) relationship (QSPR/QSAR) analysis enabling biomarker discovery, mechanistic studies, drug development, and technological evaluations in medicinal and pharmaceutical industries [23-28]. However, the use of the molecular similarity concept for QSRR modelling was seldom reported before Wang et al. [29] presented a compound classification method based on log D profile similarity, resulting in enhanced elution order prediction in acidic and basic chromatographic conditions. Muteki et al. [5] have also assessed the reliability of QSRR prediction and found that QSRR methodology based on compound classification significantly improved retention time prediction in comparison with the models derived from the whole dataset.

Previous work from our group has demonstrated [30] that application of the federation of local models strategy, which involves scanning a database to find those molecules that are most structurally similar to the test analyte and constructing a local model for each test compound based on its top ranked similar molecules, may help to improve the prediction accuracy of QSRR models. This compound-classification-based QSRR strategy successfully utilised Tanimoto [31] cluster analysis to predict retention times of studied test probes in a HILIC database using an amide column

([32], and Chapter 4). However the further application of the proposed modelling approach in datasets collected from other HILIC stationary phases or other test analytes showed varying degrees of success, possibly due to the complex retention mechanisms at play in the HILIC mode [33].

A comparison of global modelling (using the whole dataset for model production), modelling based on Tanimoto similarity (TS) clustering, and modelling based on a newly proposed retention time (t_R) similarity clustering method applied to a HILIC dataset has shown that while Tanimoto clustering shows an improvement in error compared to the global model, retention time clustering is by far the most successful method [34]. However, retention time clustering is unable to be applied to a real-life situation because the retention time of the compound under investigation is not known and so far no method has been found to successfully utilise retention time clustering as part of a dual-filtering approach.

In this study, a novel dual-filtering based QSRR modelling strategy has been applied successfully to a range of HILIC systems. The proposed dual-filtering approach involves selecting structurally similar training neighbours to a target according to calculated Tanimoto pairwise values, followed by further filtering according to t_R similarity found by utilising the correlation of molecular descriptors to retention time. The application of the proposed dual-filtering based QSRR modelling approach is illustrated by the prediction of retention time for various analytes on HILIC stationary phases, utilising a GA coupled with PLS for variable selection. By using this dual-filtering approach, reliable and accurate GA-PLS models have been established over a wide range of HILIC datasets. The performance of the GA-PLS models derived from both diverse and global datasets, and TS-based QSRR models, is also compared to dual-filtering based QSRR models. Finally, in order to obtain some insight into the HILIC mechanism,

the selected molecular descriptors in the proposed dual-filtering process have been investigated.

5.2 Method

5.2.1 Data set

Five data sets composed of 98 analytes presented in Table 5.1 with experimental retention times over five HILIC stationary phases (bare silica, amine, amide, diol and zwitterionic) were used. The isocratic eluent contained 90:10 acetonitrile−formate buffer solution. Formate buffer 100 mM was prepared with an adapted volume of ammonium formate and the pH adjusted to 3.0 with formic acid. Details of collected data are presented in Chapter 2.

The retention times observed on each stationary phase are provided in Table 5.1.

5.2.2 Model generation

5.2.2.1 Similarity searching

In the similarity searching procedure, the test analyte was matched against each compound of the database in turn, with the chosen similarity measure being used to compute the degree of resemblance in each case; the resulting set of similarity scores was then ranked in decreasing similarity order. Structural similarity searching was carried out using ChemAxon hashed fingerprint [35] descriptors and the Tanimoto similarity coefficient [31] with a threshold cutoff value of 0.5.

In the case of t_R similarity searching, the absolute value of the ratio of each training compound's retention factor (k) with the k-value of the test analyte was utilised as the t_R similarity measure with the threshold value of k-ratio < 1.5. This process was followed by construction of a predictive QSRR model on the filtered training set for each target compound.

5.2.2.2 Dual-filtering based QSRRs

A dual-filtering strategy was applied using a combination of structural similarity searching and t_R similarity searching. The scheme of the proposed method is depicted in Figure 5.1, which involves the adoption of four main steps: (A) Tanimoto similarity (TS) searching to filter compounds into classes having similar structures to the test compound (primary filter in Figure 5.1); (B) searching the nearest neighbour to the test analyte utilising a high correlated molecular descriptor to retention time ;(C) t_R similarity searching using the nearest neighbour as a reference to classify top ranked structurally similar compounds into subclasses having similar retention time (t_R) (secondary filter in Figure 5.1); (D) GA-PLS to predict the test compound's retention time based on compound properties. The method of combining TS and t_R similarity allows the identification of the most appropriate compound subclass modelling and the minimisation of the retention time prediction error.

5.3 Results and discussion

HILIC, unlike reversed-phase liquid chromatography (RPLC), has been successfully used to generate a large number of reproducible retention data of a wide range of hydrophilic compounds. In this work, such data were analysed using a dual-filtering based QSRR modelling approach to obtain local models describing the behaviour of 98 pharmaceutical compounds for various stationary phases.

In this section, first the concept of structural similarity in QSRR modelling is discussed. The results for a wide range of HILIC systems agree with previous studies [30, 34] in that there is an improvement in the predictive power of TS-based QSRR modelling compared to global

Table 5.1. Experimentally obtained retention times on the five stationary phases.

		Retention times				
nr.	Analyte	Zwitterionic	Amide	Amine	Bare silica	Diol
1	2'-Deoxyadenosine	5.06	4.73	4.96	2.57	3.52
2	2'-Deoxycytidine	10.17	10.49	11.06	4.32	3.53
3	2',3'-Dideoxyadenosine	4.00	3.98	3.85	2.34	3.77
4	2'-Deoxyguanosine	10.58	8.93	11.91	3.82	3.26
5	3'-Deoxyguanosine	9.74	8.70	11.14	3.54	3.15
6	5'-Methyluridine	4.42	4.16	5.18	1.87	2.35
7	Adenosine	5.96	5.46	6.18	2.69	3.20
8	Cytidine	12.92	12.00	15.38	4.83	3.07
9	Guanosine	13.59	10.91	16.29	4.22	3.33
10	Inosine	8.91	6.91	10.89	3.31	2.78
11	Thymidine	3.52	3.53	3.87	1.67	2.37
12	Uridine	5.11	4.30	5.95	2.03	2.35
13	Acyclovir	9.35	7.60	10.31	3.84	3.29
14	2'-Deoxyuridine	3.97	3.73	4.35	1.80	2.36
15	3'-Deoxythymidine	2.68	2.95	2.86	1.44	2.37
16	2'-Deoxyinosine	7.10	5.95	8.14	3.05	2.92
17	Adrenaline	19.23	7.75	10.82	6.73	4.89
18	Noradrenaline	26.28	9.81	14.71	7.30	4.86
19	3'-Methoxytyramine	10.30	5.98	6.17	4.86	5.46
20	Isoproterenol	12.31	6.40	7.13	4.63	4.40
21	Fenotrole	8.10	6.75	5.11	2.72	3.78
22	Terbutaline	9.13	5.88	5.71	3.53	4.06
23	Salbutamol	10.13	6.49	6.51	4.59	4.62
24	Ritudrine	5.68	5.43	3.77	2.47	4.11
25	Metaproterenol	10.96	6.41	6.63	4.10	4.23
26	Synephrine	11.49	6.32	6.84	5.20	5.17
27	Dopamine	15.78	7.64	9.23	5.62	5.05
28	N-methylephrine	6.03	3.98	3.80	3.90	5.75
29	Norphenylephrine	14.16	7.58	8.81	5.26	5.04
30	Phenylephrine	11.07	6.06	6.73	4.92	5.06
31	Tyramine	10.17	6.26	6.18	4.54	5.37
32	Normetanephrine	14.54	7.55	8.79	5.85	5.22
33	Octopamine	14.58	7.93	8.93	5.52	5.09
34	Methoxamine	6.91	5.04	4.66	3.61	5.39

35	Isoxuprine	3.92	3.98	2.87	2.11	4.25
36	Pentoxyfylline	2.16	2.37	2.19	1.32	2.60
37	Guanine	9.90	9.73	10.41	3.64	3.84
38	Xanthine	4.85	5.39	5.50	2.04	2.67
39	Caffeine	2.31	2.50	2.35	1.39	2.65
40	Theophylline	2.67	3.02	2.78	1.46	2.66
41	Theobromine	2.85	3.11	2.93	1.60	2.65
42	Diphylline	3.48	3.50	3.75	1.87	2.67
43	7-Hydroxyethyltheophylline	2.75	2.92	2.84	1.54	2.63
44	1-Methyluric acid	6.46	7.36	8.38	2.37	2.62
45	1-Methylguanine	5.98	6.18	6.15	2.86	3.70
46	9-Methylguanine	6.96	6.57	7.22	3.10	3.64
47	Uric acid	13.10	13.83	18.40	3.83	2.75
48	3,7-Dimethyluric acid	4.86	5.21	5.61	2.18	2.65
49	7-Methylxanthine	3.67	4.03	3.96	1.84	2.67
50	Hypoxanthine	5.98	5.52	6.56	2.67	3.24
51	Proxyphylline	2.52	2.75	2.61	1.45	2.58
52	1,7-Dimethyluric acid	4.56	5.27	5.67	1.99	2.62
53	1,3-Dimethyluric acid	4.13	4.72	4.75	1.92	2.60
54	1,3,7-Trimethyluric acid	3.10	2.81	3.20	1.90	2.99
55	Salicylic acid	2.96	3.70	4.75	1.34	2.28
56	5-Methylsalicylic acid	2.80	3.58	4.32	1.31	2.33
57	4-Hydroxybenzoic acid	2.59	2.75	2.83	1.24	2.22
58	3-Hydroxybenzoic acid	3.26	3.09	3.85	1.40	2.26
59	2,3-Dihydroxybenzoic acid	4.20	4.71	7.72	1.55	2.24
60	2,4-Dihydroxybenzoic acid	4.29	4.60	6.30	1.54	2.27
61	2,5-Dihydroxybenzoic acid	4.34	4.77	7.44	1.56	2.17
62	3,4-Dihydroxybenzoic acid	3.73	3.23	4.35	1.44	2.21
63	3,5-Dihydroxybenzoic acid	4.77	3.78	5.61	1.60	2.22
64	Benzoic acid	2.55	2.68	2.90	1.30	2.32
65	3-Amino-4-hydroxybenzoic acid	3.15	3.04	3.54	1.46	2.22
66	4-Aminobenzoic acid	2.32	2.47	2.42	1.53	2.21
67	4-Aminosalicylic acid	4.01	3.59	4.67	1.19	2.26
68	3-Aminobenzoic acid	3.24	2.82	3.54	1.37	2.28
69	Vanillic acid	2.70	2.76	2.95	1.29	2.25
70	Syringic acid	3.01	2.89	3.32	1.40	2.27
71	2-Methoxybenzoic acid	2.65	2.69	3.31	1.37	2.28
72	P-Toluic acid	2.26	2.39	2.48	1.21	2.33

73	4-Nitrophenyl-B-D-glycopyranoside	3.37	3.48	4.02	1.50	2.27
74	Tyrosine	36.55	24.42	45.31	13.03	6.78
75	Satolol	7.17	4.76	4.59	3.77	4.32
76	Atenolol	13.21	7.45	7.67	7.35	5.79
77	Vadarabine	7.25	6.34	7.34	3.00	3.25
78	Tryptophan	23.25	19.08	25.78	8.91	7.07
79	BTMA	6.85	3.64	3.59	5.70	7.23
80	PTMA	8.05	3.92	4.01	6.83	7.44
81	Labetalol	10.60	4.77	3.44	1.91	4.43
82	Nadolol	11.12	7.17	7.26	5.75	5.48
83	Propranolol	4.53	3.93	3.19	2.58	5.20
84	Adenine	5.58	5.32	5.24	2.80	4.75
85	Uracil	3.26	3.27	3.49	1.57	2.39
86	Thymine	2.93	3.18	3.14	1.46	2.40
87	Cytosine	9.19	10.91	9.55	4.19	4.30
88	Pindolol	5.47	4.21	3.46	2.80	4.66
89	Alprenolol	4.14	3.54	3.12	2.46	4.83
90	Nicotinic acid	12.65	8.72	19.61	5.66	4.27
91	4-Hydroxybenzenesulfonic acid	4.85	5.85	9.99	1.55	1.87
92	4-Aminophenylacetic acid	3.29	2.83	3.85	1.58	2.27
93	P-Toluenesulfonic acid	2.97	3.97	5.56	1.29	1.87
94	Tropic acid	3.64	3.25	4.88	1.70	2.34
95	2-Phenylethylamine	7.14	4.88	4.59	3.84	5.86
96	Phenylalanine	21.58	16.39	27.46	10.04	7.36
97	Mandelic acid	6.00	5.64	11.72	2.43	2.68
98	5-Sulfosalicylic acid	19.50	19.06	104.06	3.80	2.08

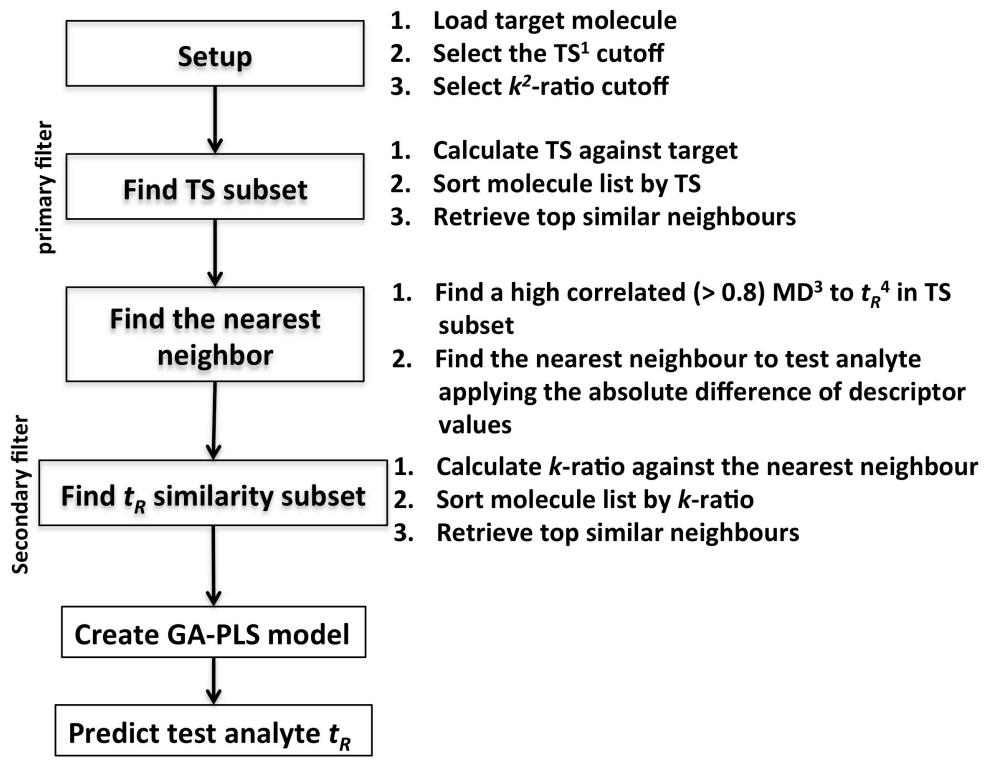

Figure 5.1. Scheme of dual-filtering based QSRR strategy in this study. [1]Tanimoto similarity, [2]retention factor, [3]molecular descriptor, and [4]retention time.

modelling, however the improvement is not sufficient to enable detailed selection of the optimal stationary and mobile phase conditions. This may be due to the lack of sufficient very similar compounds in the initial database [34]. Second, the role of t_R similarity filtering in QSRR modelling is studied, leading to excellent prediction accuracy in agreement with previous results [34]. However, retention factor filtering (referred to herein as *k*-ratio filtering) is not applicable practically as the retention time for any new compound is unknown, thus an alternative filtering tool is required which enables the identification of training set compounds having similar retention times to the retention time of the target. To overcome this drawback, a two-stage filtering approach is applied which combines the concept of structural similarity and retention time similarity into QSRR methodology and uses correlation to a molecular descriptor to search the nearest neighbour, which was used further as a reference in the secondary filtering process (Figure 5.1). The proposed QSRR model significantly improves the retention time predictability compared with diverse, global, and TS-based QSRR models. The dual-filtering based QSRR modelling is discussed in detail in Section 5.3.3. Finally, the dual-filtering strategy is investigated to provide some insight into the separation mechanisms operating in HILIC mode.

5.3.1 Tanimoto similarity (TS) searching into QSRR modelling

The Tanimoto coefficient as a measure of molecular similarity [36] was used to carry out Tanimoto Similarity (TS) searching. For a given test molecule, a subset of molecules with the Tanimoto coefficients greater than a preselected threshold was found from a search of the full database. In order to show the internal similarity of the obtained subsets with the Tanimoto threshold of 0.5, the distribution of the range of Tanimoto

pairwise values of each test analyte with all other compounds in the local training set is shown in Figure 5.2A.

A local GA-PLS model based on molecular descriptors computed using Density Functional Theory was developed for each of the test analytes using a filtered training set and the local model was compared to the global model given by whole dataset. Statistics for the retention time predictions of analytes using both local and global modelling are presented in Table 5.2 and Table 5.3 and show a total decrease of mean RMSECV (equation 2.6) from 2.61 to 0.60, a total increase of Q^2 (equation 2.7) from 0.54 to 0.89 and a total decrease of average RMSEP (equation 2.4) of more than 19% from global to local modelling over all HILIC stationary phases. The obtained results indicate that the local QSRR models greatly outperform the global models. To be able to best investigate the impact of structural similarity on QSRR prediction accuracy, the local models were also compared to diverse models given by a training set consisting of 26 dissimilar compounds with an average Tanimoto value of 0.2 (Figure 5.2B), and the corresponding RMSEP values of the obtained QSRR models are presented in Table 5.4, confirming again the positive effect of structural similarity on QSRR accuracy. The predicted retention times of test compounds applying TS-based, global and diverse QSRR modelling are presented in the Tables 5.5-5.7.

In addition, average similarity can be used as an indicator of the prediction confidence. Figure 5.3 shows a plot of mean absolute error of prediction (MAE prediction, equation 2.3) *versus* the average similarity value of the training set for all test compounds in the similar, global and diverse training sets over all studied chromatographic conditions. The higher the average similarity value in the training set, the more likely a suitable model with higher predictive ability will be obtained, in agreement

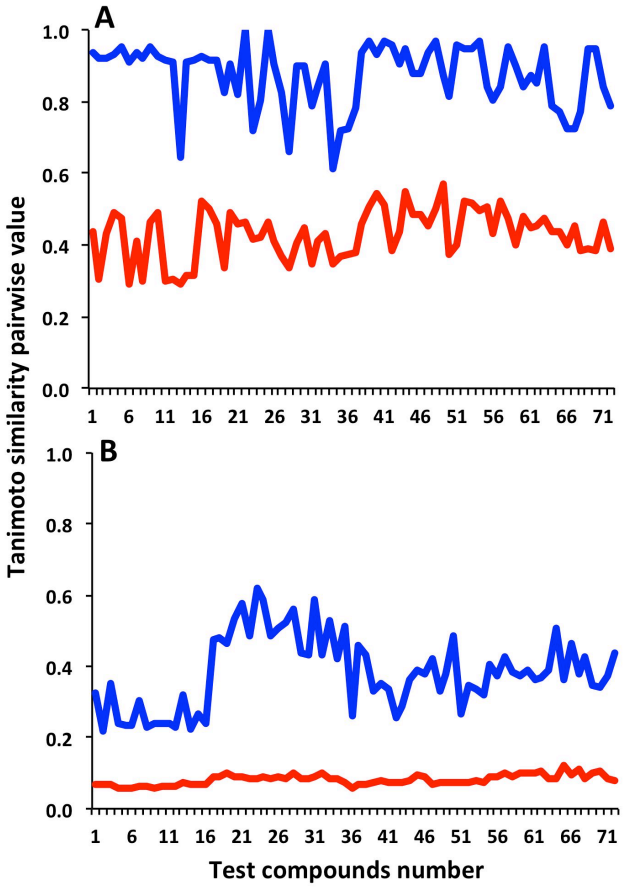

Figure 5.2. Distribution of the range of similarity for A. Tanimoto similarity based subsets, and B. diverse subsets with most similar values (royal blue lines), and least similar values (red line) in the dataset.

Table 5.2. Global and TS-based QSRR models performance summary

	RMSEP (%)	
System	Global model	TS-based model
Zwitterionic	68.46	32.22
Amide	41.95	28.37
Amine	56.17	52.04
Bare silica	49.07	18.27
Diol	17.10	4.60
Average value	46.55	27.10

Method abbreviations explained in the text.

Table 5.3. Predictive performance of TS-based and global QSRRs

	TS-based QSRR models with		global QSRR models with	
System	Q^2_{CV}	RMSECV	Q^2_{CV}	RMSECV
Zwitterionic	0.88	1.02	0.46	4.18
Amide	0.91	0.61	0.50	2.48
Amine	0.87	1.07	0.53	4.15
Bare silica	0.84	0.22	0.48	1.52
Diol	0.93	0.06	0.75	0.70

Table 5.4. Diverse QSRR models performance summary

System	RMSEP (%)	Q^2_{CV}	RMSECV (min)
Zwitterionic	83.63	0.45	3.21
Amide	120.52	0.82	1.98
Amine	67.19	0.45	2.62
Bare silica	54.45	0.48	1.01
Diol	20.6	0.82	0.41

Method abbreviations explained in the text

Table 5.5. Median predicted retention times using the TS-based QSRR modelling strategy.

		Retention times				
nr.	Analytes	Zwitterionic	Amide	Amine	Bare silica	Diol
	Nucleosides					
1	2'-Deoxyadenosine	6.78	6.06	7.34	2.93	3.47
2	2'-Deoxycytidine	6.75	6.27	7.11	3.27	2.89
3	2',3'-Dideoxyadenosine	3.66	3.32	4.54	2.49	3.62
4	2'-Deoxyguanosine	10.32	9.13	11.26	3.90	3.44
5	3'-Deoxyguanosine	10.22	7.72	11.51	3.27	3.17
6	5'-Methyluridine	5.88	5.98	7.17	2.02	2.37
7	Adenosine	9.35	8.99	11.86	3.59	3.49
8	Cytidine	6.18	5.82	6.54	2.97	3.17
9	Guanosine	9.89	8.52	11.86	3.54	2.99
10	Inosine	8.47	7.11	10.22	3.41	2.98
11	Thymidine	3.62	3.60	3.25	1.67	2.50
12	Uridine	9.36	8.54	11.61	2.93	2.34
13	Acyclovir	12.70	11.24	12.62	4.78	3.52

14	2'-Deoxyuridine	5.62	5.38	7.88	2.05	2.65
15	3'-Deoxythymidine	0.04	0.91	-0.50	0.89	2.29
16	2'-Deoxyinosine	7.74	6.00	8.06	3.19	2.69
	β-Adrenergic Agonists					
17	Adrenaline	17.05	7.71	9.82	6.01	4.88
18	Noradrenaline	20.58	9.21	12.05	6.57	4.84
19	3'-Methoxytyramine	8.05	5.89	5.30	4.79	5.52
20	Isoproterenol	15.18	6.83	8.66	5.27	4.58
21	Fenotrole	11.03	6.36	6.92	3.28	3.79
22	Terbutaline	12.04	6.71	7.27	4.35	4.15
23	Salbutamol	7.21	5.29	4.20	3.22	4.31
24	Ritudrine	3.27	6.14	3.02	2.63	4.23
25	Metaproterenol	11.37	6.47	6.87	4.34	4.29
26	Synephrine	11.88	6.34	7.25	5.19	5.18
27	Dopamine	18.04	7.63	9.99	5.88	5.03
28	N-methylephrine	4.31	3.97	3.41	4.00	5.37
29	Norphenylephrine	16.28	7.98	9.51	5.60	5.02
30	Phenylephrine	10.93	6.19	6.70	4.95	5.10
31	Tyramine	8.54	6.24	5.40	3.82	5.28
32	Normetanephrine	16.58	7.49	9.61	5.96	5.13
33	Octopamine	16.16	8.05	10.11	6.07	5.19
34	Methoxamine	8.27	6.13	5.77	4.42	5.31
35	Isoxuprine	1.44	4.78	1.98	1.90	4.35
	Xanthines and Uric Acids					
36	Pentoxyfylline	3.43	4.08	8.71	1.80	2.54
37	Guanine	7.04	7.25	7.48	2.64	3.73
38	Xanthine	9.78	10.07	12.23	3.27	2.80
39	Caffeine	0.78	1.27	0.07	0.93	2.53
40	Theophylline	2.62	2.97	2.68	1.62	2.66
41	Theobromine	2.72	2.87	2.57	1.54	2.60
42	Diphylline	5.02	4.76	5.48	2.30	2.76
43	7-Hydroxyethyltheophylline	3.45	3.66	3.66	1.58	2.58
44	1-Methyluric acid	7.48	7.87	9.52	2.49	2.71
45	1-Methylguanine	5.28	5.24	4.95	2.60	3.48
46	9-Methylguanine	7.45	6.87	8.75	3.53	3.68
47	Uric acid	6.18	7.26	7.66	2.52	2.56
48	3,7-Dimethyluric acid	5.87	6.67	8.32	2.13	2.60

49	7-Methylxanthine	5.59	5.79	6.75	2.19	2.71
50	Hypoxanthine	3.46	3.35	2.89	2.10	3.37
51	Proxyphylline	2.34	2.81	2.36	1.47	2.66
52	1,7-Dimethyluric acid	4.59	5.48	6.09	2.21	2.64
53	1,3-Dimethyluric acid	4.89	5.45	5.97	1.96	2.58
54	1,3,7-Trimethyluric acid	2.53	3.04	1.77	1.43	2.56
	Benzoic acids					
55	Salicylic acid	3.19	3.84	5.01	1.24	2.31
56	5-Methylsalicylic acid	2.93	3.45	4.48	1.38	2.30
57	4-Hydroxybenzoic acid	2.66	2.75	2.90	1.34	2.24
58	3-Hydroxybenzoic acid	3.54	3.08	4.19	1.43	2.22
59	2,3-Dihydroxybenzoic acid	4.17	4.73	7.24	1.54	2.22
60	2,4-Dihydroxybenzoic acid	4.01	4.00	6.11	1.37	2.21
61	2,5-Dihydroxybenzoic acid	4.17	4.59	7.55	1.60	2.24
62	3,4-Dihydroxybenzoic acid	3.52	3.33	4.17	1.46	2.19
63	3,5-Dihydroxybenzoic acid	4.42	3.73	5.31	1.40	2.23
64	Benzoic acid	2.21	2.45	2.71	1.36	2.27
65	3-Amino-4-hydroxybenzoic acid	3.58	3.09	4.14	1.29	2.24
66	4-Aminobenzoic acid	2.56	2.67	2.50	1.23	2.27
67	4-Aminosalicylic acid	3.50	3.38	4.16	1.54	2.23
68	3-Aminobenzoic acid	2.70	2.73	3.11	1.38	2.29
69	Vanillic acid	3.03	2.64	3.05	1.41	2.25
70	Syringic acid	2.85	2.92	3.82	1.46	2.25
71	2-Methoxybenzoic acid	2.83	2.83	3.03	1.29	2.32
72	P-Toluic acid	2.36	2.55	2.77	1.37	2.28

Table 5.6. Median predicted retention times using the global QSRR modelling strategy.

		Retention times in				
nr.	Analytes	Zwitterionic	Amide	Amine	Bare silica	Diol
	Nucleosides					
1	2'-Deoxyadenosine	6.66	6.63	7.76	3.86	4.02

2	2'-Deoxycytidine	8.24	7.44	9.02	3.80	3.34
3	2',3'-Dideoxyadenosine	8.67	6.73	7.99	4.41	4.54
4	2'-Deoxyguanosine	8.47	7.29	7.23	2.72	2.91
5	3'-Deoxyguanosine	7.11	6.43	7.32	2.79	2.99
6	5'-Methyluridine	9.05	5.70	7.42	3.10	2.50
7	Adenosine	10.73	6.74	7.83	3.86	3.74
8	Cytidine	7.72	5.79	7.00	3.04	2.89
9	Guanosine	11.24	8.28	11.09	3.07	2.79
10	Inosine	8.44	6.59	7.71	2.84	2.82
11	Thymidine	4.26	4.21	4.35	2.93	2.56
12	Uridine	10.69	7.03	10.55	3.56	2.41
13	Acyclovir	10.15	8.30	12.19	3.68	3.62
14	2'-Deoxyuridine	5.56	5.04	7.05	2.46	2.57
15	3'-Deoxythymidine	3.58	4.18	5.16	2.66	2.86
16	2'-Deoxyinosine	4.99	4.88	6.08	2.45	2.80
	β-Adrenergic Agonists					
17	Adrenaline	15.90	8.54	10.73	5.93	5.02
18	Noradrenaline	17.08	9.30	12.58	6.17	4.88
19	3'-Methoxytyramine	11.72	5.59	6.83	4.84	5.18
20	Isoproterenol	13.56	7.55	9.66	4.96	4.74
21	Fenotrole	10.14	6.43	6.09	2.43	4.45
22	Terbutaline	12.04	7.16	6.23	5.04	5.06
23	Salbutamol	8.92	5.16	4.16	4.04	4.71
24	Ritudrine	10.07	5.86	5.39	3.93	4.65
25	Metaproterenol	10.85	6.73	6.67	4.58	4.70
26	Synephrine	14.40	8.23	10.78	6.35	5.48
27	Dopamine	14.86	7.13	9.14	5.04	4.96
28	N-methylephrine	9.47	5.74	6.81	5.34	5.63
29	Norphenylephrine	12.74	6.82	8.90	4.74	4.99
30	Phenylephrine	12.22	6.35	8.02	5.40	5.20
31	Tyramine	12.11	6.76	7.03	4.53	5.48
32	Normetanephrine	14.01	6.90	8.20	4.97	4.87
33	Octopamine	14.67	7.86	9.58	5.47	5.19
34	Methoxamine	8.54	3.65	2.24	3.35	4.62
35	Isoxuprine	5.53	3.93	1.23	2.02	4.61
	Xanthines and Uric Acids					
36	Pentoxyfylline	4.89	5.66	8.86	3.43	3.92
37	Guanine	8.38	6.86	8.70	3.28	3.70

38	Xanthine	9.55	8.46	11.70	3.57	3.04
39	Caffeine	0.51	0.33	-0.48	1.01	2.86
40	Theophylline	3.82	2.88	3.33	2.39	3.18
41	Theobromine	1.99	2.74	2.44	1.75	2.83
42	Diphylline	5.78	4.36	5.32	2.55	2.45
43	7-Hydroxyethyltheophylline	2.18	2.70	2.08	2.17	2.79
44	1-Methyluric acid	7.55	8.04	9.60	2.88	2.67
45	1-Methylguanine	7.29	6.15	5.86	3.09	3.77
46	9-Methylguanine	9.06	8.48	10.63	3.96	4.19
47	Uric acid	7.95	8.12	9.53	2.95	2.87
48	3,7-Dimethyluric acid	4.30	4.90	5.94	1.51	2.58
49	7-Methylxanthine	5.25	6.09	6.04	2.18	2.86
50	Hypoxanthine	6.18	5.45	7.67	2.50	3.30
51	Proxyphylline	2.93	2.88	3.04	2.36	2.83
52	1,7-Dimethyluric acid	5.73	5.63	6.16	2.17	2.78
53	1,3-Dimethyluric acid	4.41	4.83	6.55	1.96	2.53
54	1,3,7-Trimethyluric acid	2.37	3.48	2.84	1.28	2.70
	Benzoic acids					
55	Salicylic acid	3.69	4.94	8.87	0.72	2.26
56	5-Methylsalicylic acid	4.47	4.39	7.48	1.50	2.13
57	4-Hydroxybenzoic acid	3.29	3.08	4.26	1.17	1.68
58	3-Hydroxybenzoic acid	1.52	1.74	1.41	1.04	1.61
59	2,3-Dihydroxybenzoic acid	5.20	4.71	8.24	1.63	2.22
60	2,4-Dihydroxybenzoic acid	3.01	3.00	3.00	0.42	1.57
61	2,5-Dihydroxybenzoic acid	4.09	4.85	7.18	1.09	2.01
62	3,4-Dihydroxybenzoic acid	4.91	3.26	4.47	1.48	1.32
63	3,5-Dihydroxybenzoic acid	0.83	2.29	1.88	-0.30	1.19
64	Benzoic acid	1.86	2.03	2.89	1.45	2.59
65	3-Amino-4-hydroxybenzoic acid	13.39	7.54	10.39	3.95	3.27
66	4-Aminobenzoic acid	3.59	2.97	3.50	1.56	2.80
67	4-Aminosalicylic acid	3.36	3.13	2.91	1.06	2.20
68	3-Aminobenzoic acid	9.76	5.92	7.82	3.59	3.58
69	Vanillic acid	3.88	3.44	5.27	1.64	2.26

70	Syringic acid	-0.05	0.59	-0.51	-0.48	1.85
71	2-Methoxybenzoic acid	-2.81	-0.44	-1.97	-0.52	2.07
72	P-Toluic acid	1.47	2.08	3.06	1.50	2.72

Table 5.7. Median predicted retention times using the diverse QSRR models.

nr.	Analytes	Retention times in				
		Zwitterionic	Amide	Amine	Bare silica	Diol
	Nucleosides					
1	2'-Deoxyadenosine	8.56	6.77	4.80	2.96	3.37
2	2'-Deoxycytidine	5.49	8.29	6.25	1.51	2.73
3	2',3'-Dideoxyadenosine	8.38	11.11	5.52	2.94	3.96
4	2'-Deoxyguanosine	6.29	8.34	5.08	2.48	2.92
5	3'-Deoxyguanosine	2.30	6.04	4.86	1.93	2.76
6	5'-Methyluridine	10.01	7.24	5.81	2.01	2.55
7	Adenosine	5.23	6.75	6.47	3.00	3.19
8	Cytidine	6.82	7.08	7.29	1.58	2.57
9	Guanosine	5.63	6.57	6.64	2.59	2.46
10	Inosine	5.17	6.82	6.12	1.46	2.44
11	Thymidine	6.73	8.21	4.46	1.10	2.46
12	Uridine	7.28	5.06	6.46	1.64	1.97
13	Acyclovir	6.59	10.13	6.86	2.41	2.97
14	2'-Deoxyuridine	4.66	10.04	5.60	1.57	2.11
15	3'-Deoxythymidine	5.87	8.62	4.97	2.04	2.65
16	2'-Deoxyinosine	5.77	10.15	5.68	2.28	2.82
	β-Adrenergic Agonists					
17	Adrenaline	6.44	4.65	8.00	3.61	3.91
18	Noradrenaline	5.68	4.73	8.06	3.61	3.74
19	3'-Methoxytyramine	8.71	4.50	5.55	3.93	4.05
20	Isoproterenol	8.10	4.81	8.12	4.14	4.31
21	Fenotrole	9.34	5.83	4.66	2.91	4.17
22	Terbutaline	10.10	5.62	6.24	6.03	4.85
23	Salbutamol	8.17	5.35	5.57	5.72	4.00
24	Ritudrine	11.23	5.55	5.02	4.00	4.74
25	Metaproterenol	8.15	5.05	5.66	4.89	4.65
26	Synephrine	5.54	4.77	5.84	3.50	4.61
27	Dopamine	6.86	4.91	7.84	3.67	4.32
28	N-methylephrine	5.84	3.72	4.13	4.38	4.28

29	Norphenylephrine	5.15	4.70	7.28	3.61	4.14
30	Phenylephrine	4.56	4.36	6.46	3.77	4.10
31	Tyramine	6.64	4.66	5.06	3.91	5.06
32	Normetanephrine	5.79	4.92	6.16	3.93	3.54
33	Octopamine	4.44	4.59	6.19	3.28	3.82
34	Methoxamine	6.01	3.83	3.59	4.89	4.07
35	Isoxuprine	7.64	3.31	1.81	3.43	4.62
	Xanthines and Uric Acids					
36	Pentoxyfylline	7.58	12.65	4.02	2.97	3.70
37	Guanine	4.69	6.16	8.43	2.52	2.64
38	Xanthine	4.03	5.13	7.09	3.08	3.08
39	Caffeine	7.33	9.62	4.97	3.16	3.38
40	Theophylline	8.18	10.22	6.16	3.90	3.58
41	Theobromine	5.47	8.72	5.61	2.17	3.04
42	Diphylline	8.00	6.26	5.70	2.68	3.16
43	7-Hydroxyethyltheophylline	8.79	9.28	5.01	3.21	3.35
44	1-Methyluric acid	5.15	11.74	6.94	2.29	2.14
45	1-Methylguanine	3.20	9.87	7.03	2.23	3.02
46	9-Methylguanine	7.73	12.82	7.79	2.87	3.52
47	Uric acid	7.02	10.08	8.45	2.91	2.83
48	3,7-Dimethyluric acid	1.08	10.33	5.33	2.47	2.92
49	7-Methylxanthine	3.99	10.15	6.69	2.38	3.01
50	Hypoxanthine	6.75	8.25	9.28	3.91	4.23
51	Proxyphylline	8.34	11.04	3.98	2.43	3.25
52	1,7-Dimethyluric acid	1.86	8.21	6.21	2.53	2.99
53	1,3-Dimethyluric acid	8.92	8.34	5.54	3.02	3.24
54	1,3,7-Trimethyluric acid	7.24	10.43	4.83	2.42	3.23
	Benzoic acids					
55	Salicylic acid	3.18	6.02	6.92	1.54	2.53
56	5-Methylsalicylic acid	3.59	6.00	7.08	1.62	2.43
57	4-Hydroxybenzoic acid	4.57	5.91	8.99	1.96	2.51
58	3-Hydroxybenzoic acid	4.73	6.37	6.04	2.04	2.55
59	2,3-Dihydroxybenzoic acid	3.50	5.78	7.19	1.84	2.74
60	2,4-Dihydroxybenzoic acid	4.23	6.86	5.50	1.95	2.71
61	2,5-Dihydroxybenzoic acid	3.38	6.99	4.59	1.82	2.47

62	3,4-Dihydroxybenzoic acid	4.80	5.92	9.00	2.04	2.56
63	3,5-Dihydroxybenzoic acid	4.90	3.23	6.32	1.94	2.28
64	Benzoic acid	4.54	6.98	6.66	2.03	2.71
65	3-Amino-4-hydroxybenzoic acid	4.21	9.96	4.99	2.01	2.28
66	4-Aminobenzoic acid	4.60	6.58	6.13	1.90	2.34
67	4-Aminosalicylic acid	4.08	3.41	5.00	1.91	2.26
68	3-Aminobenzoic acid	4.50	6.84	5.44	1.90	2.43
69	Vanillic acid	4.06	5.12	5.26	2.01	2.63
70	Syringic acid	3.40	6.18	3.03	2.13	2.59
71	2-Methoxybenzoic acid	3.96	5.10	3.51	1.97	2.63
72	P-Toluic acid	4.49	8.15	1.61	1.94	2.83

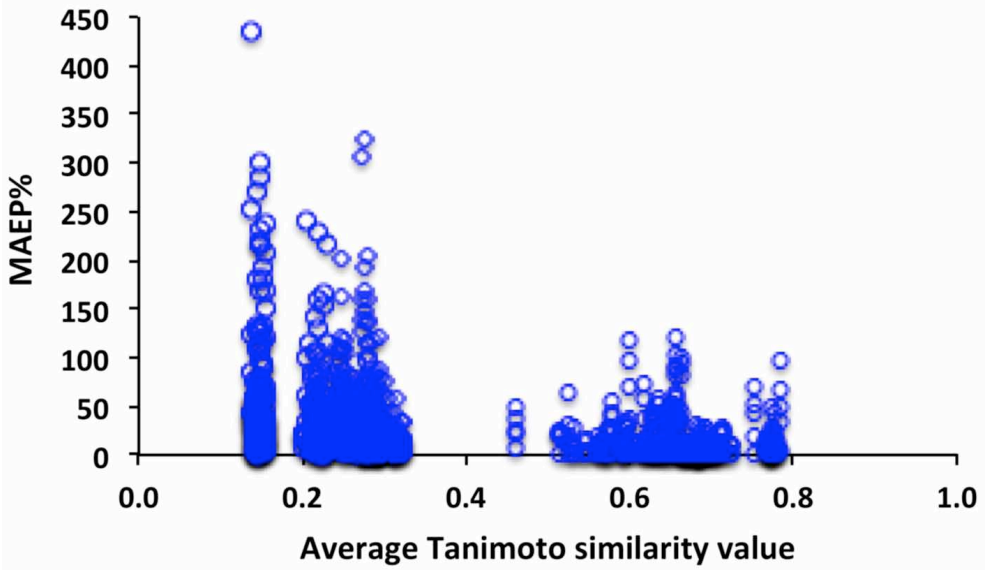

Figure 5.3. MAEP (%) vs. average similarity of compounds in the training set to each test analyte applying global, diverse and TS-based QSRR models over all HILIC systems. A total of 1077 data points included on the graph.

with previous work [34]. This is seen from the fact that the spread of MAE values decreases as average similarity increases. At an average similarity > 0.5, the MAEP values are lower and less variable than when the average similarity was < 0.4.

Although these results show lower RMSEP values (Table 5.2) for test analytes applying a TS-based QSRR modelling strategy compared with traditional QSRR models without compound filtering, the prediction accuracy achieved is generally insufficient to support detailed HILIC method development.

5.3.2 Incorporation of retention time (t_R) similarity searching into QSRR modelling

The federation of local models strategy (which involves scanning a database to find those molecules that are most structurally similar to the test analyte and generating a local model for each test compound based on its top ranked similar molecules) has been evaluated previously using retention time (t_R) similarity searching in the generation of QSRR models using a single HILIC stationary phase [34]. This approach was extended in the present study for a wider range of HILIC systems. The retention factor ratio (*k-ratio*) was used as a measure of retention time similarity with the threshold cut-off value set at 1.5. The aim of t_R similarity-based QSRR modelling was to obtain predictions with the highest achievable accuracies by establishing a training set having the closest possible chromatographic similarity (as reflected by retention factor) to the test compound. The predicted retention time of test analytes and a summary of the model validation are shown in Tables 5.8 and 5.9, respectively. The correlation of the local QSRR models obtained based on the top ranked *k-ratio* similarity compounds is presented in Figure 5.4 with an average RMSEP value of 5.46% being obtained over all compounds and all HILIC systems, which

indicates that the predicted retention values are generally in good agreement with the experimental data. However, while the t_R similarity-based QSRR modelling strategy is successful, it cannot be applied in practice for unknown analytes with unknown retention data. To overcome this drawback, a two-stage filtering approach that combines both Tanimoto similarity and t_R similarity searching tools was applied.

5.3.3 Dual-filtering based QSRR modelling

The proposed dual-filtering strategy involves TS searching, followed by t_R similarity searching using the nearest neighbour in the training set, instead of the test analyte itself. The nearest neighbour was selected based on correlations between molecular descriptors and retention time (Figure 5.1). The rationale was to first find the most structurally similar subset for the test analyte based on TS as the primary filter and to then identify the subset of compounds having similar retention times and to remove any compound having a retention time significantly different from the test analyte.

In a foundational work from this group [34], two strategies to employ t_R similarity as a secondary filter were explored. In the first approach, the *k*-ratio was calculated using the most structurally similar compound in the training set, instead of the test analyte itself. In the second approach, t_R-outliers were removed from the top ranked structurally similar compounds. These dual filters resulted in a failure to capture compounds having similar retention values to the target in the final filtered training set and therefore failed to significantly improve retention time prediction.

In the present study, to be able to employ t_R similarity as the secondary filter, the correlation between molecular descriptors and the retention times of the top ranked similar compounds for each test analyte was established

Table 5.8. Median predicted retention times using tR-similarity-based QSRR models.

nr.	Analytes	Retention times in				
		Zwitterionic	Amide	Amine	Bare silica	Diol
	Nucleosides					
1	2'-Deoxyadenosine	4.86	5.15	2.63	3.63	5.44
2	2'-Deoxycytidine	10.71	10.73	4.55	3.38	9.94
3	2',3'-Dideoxyadenosine	4.02	3.75	2.61	3.60	4.04
4	2'-Deoxyguanosine	10.92	-[a]	3.76	3.47	10.53
5	3'-Deoxyguanosine	8.65	11.25	3.32	3.24	10.32
6	5'-Methyluridine	4.24	5.15	1.96	2.36	4.17
7	Adenosine	5.37	5.98	2.59	3.12	5.28
8	Cytidine	11.10	11.36	4.63	3.12	14.08
9	Guanosine	10.12	11.80	4.38	3.10	12.80
10	Inosine	6.73	11.30	3.27	2.93	10.01
11	Thymidine	3.59	3.87	1.80	2.37	3.42
12	Uridine	4.17	5.85	1.90	2.36	4.66
13	Acyclovir	7.05	10.89	3.83	3.40	9.74
14	2'-Deoxyuridine	3.55	4.45	1.77	2.36	3.94
15	3'-Deoxythymidine	2.89	3.07	1.49	2.38	2.76
16	2'-Deoxyinosine	6.13	7.50	3.14	3.08	6.91
	β-Adrenergic Agonists					
17	Adrenaline	7.57	10.83	6.61	4.89	18.55
18	Noradrenaline	8.60	15.54	6.03	4.97	21.50
19	3'-Methoxytyramine	6.25	6.55	4.54	5.38	10.29
20	Isoproterenol	6.40	6.80	4.83	4.36	10.90
21	Fenotrole	6.41	5.12	2.77	4.07	7.31
22	Terbutaline	6.30	5.94	3.40	4.27	10.41
23	Salbutamol	6.33	6.38	4.22	4.40	10.80
24	Ritudrine	5.32	3.70	2.72	4.01	5.57
25	Metaproterenol	6.50	6.75	4.03	4.19	10.66
26	Synephrine	6.05	6.69	5.26	5.14	11.37
27	Dopamine	7.85	10.04	5.34	5.03	15.65
28	N-methylephrine	4.04	3.94	3.99	5.70	6.10
29	Norphenylephrine	7.72	8.76	5.31	5.02	14.99
30	Phenylephrine	6.24	6.91	5.00	5.11	10.88
31	Tyramine	6.01	5.96	5.01	5.08	11.27

32	Normetanephrine	7.71	8.45	5.77	5.10	14.28
33	Octopamine	7.87	9.04	5.56	5.15	14.93
34	Methoxamine	4.79	4.61	3.67	5.28	6.83
35	Isoxuprine	4.01	3.02	1.78	4.64	4.21
	Xanthines and Uric Acids					
36	Pentoxyfylline	-[a]	-[a]	1.40	2.58	-[a]
37	Guanine	9.41	11.03	3.90	3.66	9.35
38	Xanthine	5.44	5.54	2.17	2.67	4.95
39	Caffeine	2.90	-[a]	1.45	2.64	2.54
40	Theophylline	3.07	2.96	1.39	2.64	2.84
41	Theobromine	3.05	2.90	1.55	2.67	2.79
42	Diphylline	3.51	3.84	1.94	2.61	3.54
43	7-Hydroxyethyltheophylline	3.02	2.63	1.58	2.65	2.62
44	1-Methyluric acid	7.71	8.87	2.22	2.61	6.43
45	1-Methylguanine	5.74	6.24	2.93	3.74	6.26
46	9-Methylguanine	6.07	7.15	3.10	3.70	5.61
47	Uric acid	12.02	17.04	3.70	2.67	13.30
48	3,7-Dimethyluric acid	5.29	5.83	2.02	2.62	4.51
49	7-Methylxanthine	4.14	3.96	1.85	2.62	4.09
50	Hypoxanthine	5.37	6.02	2.84	3.64	5.68
51	Proxyphylline	2.89	2.90	1.43	2.61	2.75
52	1,7-Dimethyluric acid	5.17	5.68	1.89	2.63	4.82
53	1,3-Dimethyluric acid	5.17	4.81	1.89	2.63	4.18
54	1,3,7-Trimethyluric acid	2.86	3.16	1.92	3.48	2.69
	Benzoic acids					
55	Salicylic acid	3.72	4.81	1.31	2.26	2.94
56	5-Methylsalicylic acid	3.62	4.16	1.39	2.33	2.78
57	4-Hydroxybenzoic acid	2.75	2.88	1.27	-[a]	2.67
58	3-Hydroxybenzoic acid	3.26	3.98	1.41	2.27	3.28
59	2,3-Dihydroxybenzoic acid	4.72	7.39	1.55	-[a]	4.40
60	2,4-Dihydroxybenzoic acid	4.77	6.52	1.58	2.26	4.06
61	2,5-Dihydroxybenzoic acid	4.88	7.80	1.58	-[a]	4.30
62	3,4-Dihydroxybenzoic acid	3.12	4.47	1.40	-[a]	3.69
63	3,5-Dihydroxybenzoic acid	3.68	5.91	1.56	-[a]	4.13

64	Benzoic acid	2.67	2.91	1.35	2.34	2.59
65	3-Amino-4-hydroxybenzoic acid	3.16	3.46	1.48	-[a]	3.11
66	4-Aminobenzoic acid	2.74	-[a]	1.53	-[a]	2.58
67	4-Aminosalicylic acid	3.69	4.53	-[a]	2.26	3.88
68	3-Aminobenzoic acid	2.74	3.58	1.37	2.27	3.25
69	Vanillic acid	2.78	2.86	1.34	-[a]	2.63
70	Syringic acid	2.67	3.27	1.41	2.28	3.12
71	2-Methoxybenzoic acid	2.69	3.31	1.38	2.28	2.67
72	P-Toluic acid	-[a]	2.41	1.31	2.33	-[a]

[a]outliers corresponding to *k*-ratio more than 1.5.

Table 5.9. Summary of tR-similarity-based QSRRs.

System	Q^2_{CV}	RMSECV	RMSEP%
Zwitterionic	0.89	0.12	5.04
Amide	0.90	0.05	3.78
Amine	0.87	0.09	3.93
Bare silica	0.83	0.07	3.88
Diol	0.93	0.00	2.38

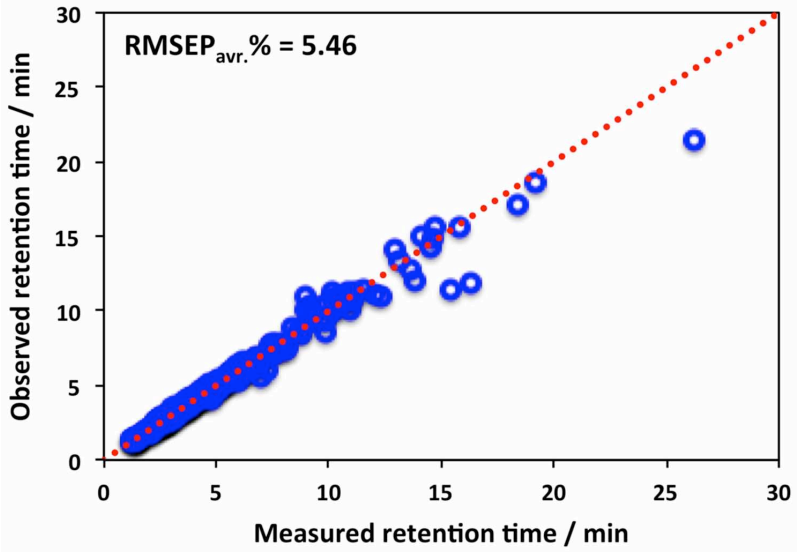

Figure 5.4. Predicted retention times vs. observed retention times of the k*ratio* filtering based QSRR models over all HILIC systems. $RMSEP_{avr.}$ is the average value of RMSEP of test compounds over five HILIC stationary phases. A total of 343 data points included on the graph.

and a molecular descriptor showing a correlation value more than 0.8 was selected and used as a measure of t_R similarity. This chosen descriptor was then used to rank the primary filter compounds and to select the nearest neighbour having lowest absolute difference descriptor values to the test analyte. The nearest neighbour was further used as a reference for t_R similarity searching applying a *k*-ratio cutoff value of 1.5. QSRR-models were then derived using the dual-filtered-training set for the selected target. The scheme of dual-filtering for 2'-deoxycytidine as an example is depicted in Figure 5.5. Here, the test analyte was 2'-deoxycytidine and the primary TS filter identified 15 compounds in the dataset which showed a TS score of ≥0.5. The retention times of these 15 compounds were strongly correlated to the HOMT (Harmonic Oscillator Model of Aromaticity index total) descriptor and this descriptor was then used to further rank the 15 compounds to identify the nearest neighbour, compound 8, having lowest absolute difference of HOMT values to the test analye. The selected nearest neighbour was then used as a reference for further filtering of the database, leading to selection of a training set of 6 compounds with *k*-ratio values of less than 1.5. These 6 compounds were then used for subsequent GA-PLS modelling.

The molecular descriptors utilised in this dual-filtering process for the various compound classes and HILIC stationary phases are shown in Tables 5.10 and Table 5.11. The potential relevance of these descriptors in understanding the HILIC experimental system is discussed in the following section.

The correlation data for the GA-PLS models obtained for dual filtered training sets for the five HILIC stationary phases are presented in Table 5.12. The internal validation by root mean squared error (RMSECV) and Q^2 gave values of 0.03-0.34 and 0.89-0.96, respectively, over all the HILIC

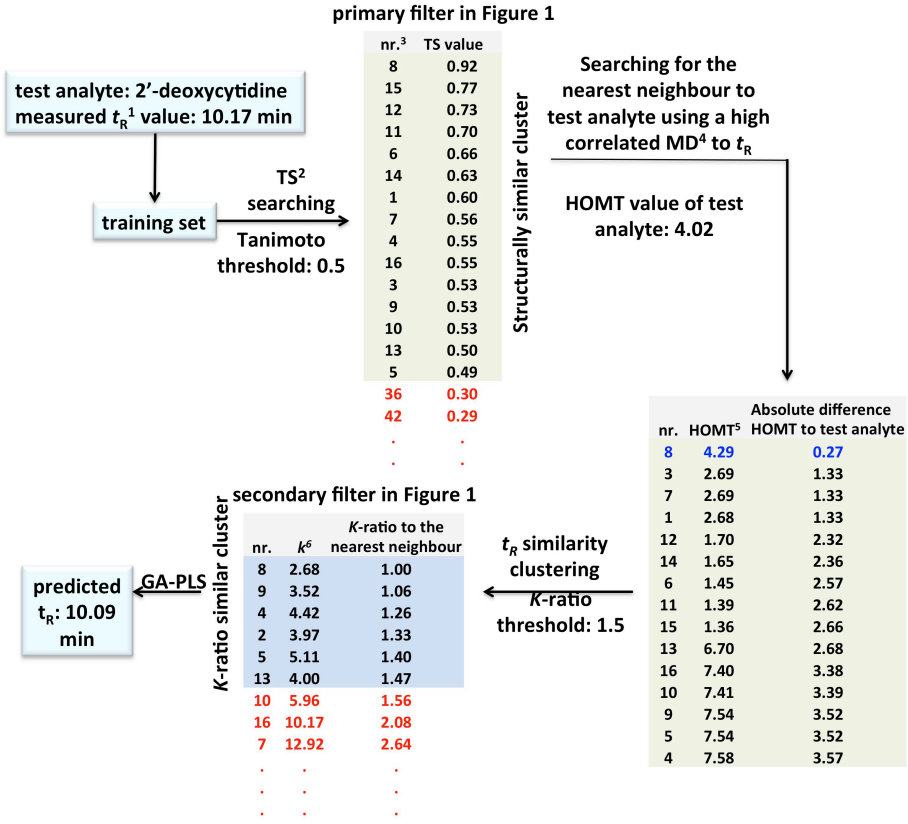

Figure 5.5. An example of dual-filtering based QSRR strategy in this study. [1]retention time, [2]Tanimoto similarity, [3]numbering of compounds in training set indicated in Table 5.1, [4]retention factor, and [5]selected high correlated molecular descriptor. The definition of the selected molecular descriptor is available in Table 5.11.

Table 5.10. Molecular descriptors in the dual-filtering strategy In the table, it is observed that the same molecular descriptor was obtained for the test compounds from the same chemical classification. Definition of molecular descriptors are available in Table 5.11.

	Stationary Phase:				
Test analytes	Diol	Bare silica	Amine	Amide	Zwitterionic
β- Agonists	Hy	MLOGP	MLOGP	Hy	MLOGP
Nucleosides	HOMT	HOMT	HOMT	HOMT	HOMT
Benzoic acids	ATSC4s	HATS2p	ATSC4s	ATSC4s	ATSC1s
Xanthines and Uric Acids	MATS1v	MATS1v	MATS1v	MATS1v	MATS1v

Table 5.11. Highly correlated molecular descriptors utilised for the dual-filtering strategy.

Molecular Descriptor	Description	Category
MLOGP	Moriguchi octanol-water partition coeff. (logP)	Molecular properties
HOMT	HOMA (Harmonic Oscillator Model of Aromaticity index) total	Geometrical descriptors
HATS2p	everage-weighted autocorrelation of lag 2 / weighted by polarisability	GETAWAY descriptors
MATS1v	Moran autocorrelation of lag 1 weighted by van der Waals volume	2D autocorrelations
ATSC4s	Centred Broto-Moreau autocorrelation of lag 4 weighted by I-state	2D autocorrelations
Hy	hydrophilic factor	Molecular properties
ATSC1s	Centred Broto-Moreau autocorrelation of lag 1 weighted by I-state	2D autocorrelations

Table 5.12. Performance summary of the dual-filtering based QSRR models

System	Q^2_{CV}	RMSECV	RMSEP (%)
Zwitterionic	0.96	0.34	10.21
Amide	0.91	0.11	13.83
Amine	0.93	0.16	11.83
Bare silica	0.89	0.17	13.75
Diol	0.94	0.03	2.95

Method abbreviations explained in Chapter 2.

Table 5.13. Median predicted retention times using the dual-filtering based QSRR modelling strategy.

		Retention times				
nr.	Analytes	Zwitterionic	Amide	Amine	Bare silica	Diol
	Nucleosides					
1	2'-Deoxyadenosine	4.71	4.14	a	a	3.55
2	2'-Deoxycytidine	10.09	10.29	11.13	4.31	3.35
3	2',3'-Dideoxyadenosine	a	4.46	4.35	a	3.59
4	2'-Deoxyguanosine	10.58	9.80	10.86	3.80	3.28
5	3'-Deoxyguanosine	10.34	9.56	12.00	3.68	3.24
6	5'-Methyluridine	4.56	4.15	a	a	2.36

7	Adenosine	a	5.09	a	a	3.16
8	Cytidine	12.85	12.06	15.48	4.76	3.04
9	Guanosine	11.66	8.38	a	3.78	3.17
10	Inosine	7.21	a	8.18	3.17	3.05
11	Thymidine	a	a	a	a	2.36
12	Uridine	4.79	4.31	5.49	a	2.16
13	Acyclovir	8.65	a	8.97	3.32	3.49
14	2'-Deoxyuridine	3.85	3.73	a	a	2.36
15	3'-Deoxythymidine	a	3.49	a	a	2.37
16	2'-Deoxyinosine	9.42	7.73	a	3.55	2.82
	β-Adrenergic Agonists					
17	Adrenaline	a	7.47	8.37	5.74	5.03
18	Noradrenaline	a	7.71	a	a	4.86
19	3'-Methoxytyramine	9.75	6.44	6.11	4.67	5.47
20	Isoproterenol	11.58	6.29	7.66	4.89	4.54
21	Fenotrole	a	7.57	6.29	3.55	3.74
22	Terbutaline	10.47	6.52	6.07	3.47	4.16
23	Salbutamol	9.99	6.33	6.45	4.25	4.75
24	Ritudrine	a	6.98	a	a	4.05
25	Metaproterenol	10.75	6.22	6.85	4.11	4.15
26	Synephrine	11.85	6.33	7.27	5.20	5.12
27	Dopamine	13.62	7.73	9.30	5.37	5.11
28	N-methylephrine	a	a	a	a	5.77
29	Norphenylephrine	14.05	7.79	8.74	5.55	5.03
30	Phenylephrine	11.09	6.25	6.82	5.11	5.12
31	Tyramine	10.19	6.757	6.41	4.84	5.26
32	Normetanephrine	a	7.74	a	6.57	5.12
33	Octopamine	14.23	7.62	6.69	5.83	5.11
34	Methoxamine	a	5.95	a	4.18	5.44
35	Isoxuprine	a	a	a	a	4.48
	Xanthines and Uric Acids					
36	Pentoxyfylline	2.13	2.29	a	a	2.64
37	Guanine	a	a	a	a	3.83
38	Xanthine	a	5.40	5.83	3.35	2.81
39	Caffeine	2.32	2.91	5.81	1.47	2.66
40	Theophylline	a	3.09	2.75	1.57	2.70
41	Theobromine	2.79	2.85	2.63	1.42	2.67
42	Diphylline	3.37	5.11	a	1.77	2.67

43	7-Hydroxyethyltheophylline	a	3.06	2.78	1.66	2.69
44	1-Methyluric acid	4.81	5.25	8.71	2.28	2.66
45	1-Methylguanine	5.84	a	5.57	2.29	3.45
46	9-Methylguanine	6.60	5.87	7.27	2.09	3.68
47	Uric acid	a	a	a	a	2.65
48	3,7-Dimethyluric acid	4.71	5.27	5.46	2.04	2.56
49	7-Methylxanthine	4.00	5.40	5.86	2.03	2.72
50	Hypoxanthine	5.33	a	a	2.31	3.40
51	Proxyphylline	a	a	a	a	2.64
52	1,7-Dimethyluric acid	5.06	4.71	5.26	2.38	2.66
53	1,3-Dimethyluric acid	5.09	5.25	a	2.18	2.52
54	1,3,7-Trimethyluric acid	3.11	2.43	a	1.68	2.67
	Benzoic acids					
55	Salicylic acid	3.05	3.42	4.93	1.30	2.25
56	5-Methylsalicylic acid	3.18	3.82	3.83	a	2.31
57	4-Hydroxybenzoic acid	2.67	2.77	2.83	1.28	2.26
58	3-Hydroxybenzoic acid	3.74	2.74	4.18	1.40	2.26
59	2,3-Dihydroxybenzoic acid	4.01	a	a	1.49	2.26
60	2,4-Dihydroxybenzoic acid	4.13	a	a	1.59	2.23
61	2,5-Dihydroxybenzoic acid	4.43	a	a	1.56	2.23
62	3,4-Dihydroxybenzoic acid	3.89	3.83	4.32	1.51	2.17
63	3,5-Dihydroxybenzoic acid	4.60	3.26	5.17	1.52	2.23
64	Benzoic acid	2.51	a	a	a	2.27
65	3-Amino-4-hydroxybenzoic acid	2.55	2.85	3.23	1.44	2.25
66	4-Aminobenzoic acid	2.57	2.65	a	1.36	2.27
67	4-Aminosalicylic acid	4.05	3.91	4.76	1.28	2.23
68	3-Aminobenzoic acid	2.45	2.47	3.70	1.34	2.28
69	Vanillic acid	2.96	3.01	3.10	1.37	2.25
70	Syringic acid	2.81	2.57	3.53	1.71	2.23
71	2-Methoxybenzoic acid	2.66	a	3.38	a	2.31
72	P-Toluic acid	2.42	2.77	a	1.32	2.27

[a] outliers corresponded to high *k*-ratio values (*k*-ratio more than 1.5)

systems, which indicates that the predicted retention values are in relatively good agreement with the experimental data. The predicted retention time of the test analytes found by applying dual-filtering based QSRR modelling is presented in Table 5.13.

An external validation was performed to evaluate the prediction accuracy of the QSRR models for the retention of the unknown compounds. A comparison between the predicted retention times and those observed experimentally for each test compound using the dual-filtering based GA-PLS model with the corresponding average RMSEP values (RMSEP values obtained for each stationary phase are presented in Table 5.12), is depicted in Figure 5.6, and is compared to the diverse, global and TS-based QSRR models in the same figure. As seen in Figure 5.6, the dual-filtering based GA-PLS models generated for all five columns were very well correlated with the experimental data, presenting an average RMSEP value of 11.01%, whereas the global and TS-based GA-PLS yielded models with average RMSEP values of 46.55% and 27.10%, respectively, while an even higher average RMSEP value of 69.28% was found for the diverse QSRR models around over all HILIC systems (Figure 5.6A). The same trend is seen in Figure 5.7 with a comparison of the predictive ability of global, TS-based and dual-filtering-based GA-PLS models on each HILIC stationary phase.

The obtained results imply that when predicting the retention time of a new compound, it is best to assign the most appropriate subset of compounds as the training set and to create the QSRR model specific to that subset, providing an acceptable level of accuracy of retention time prediction.

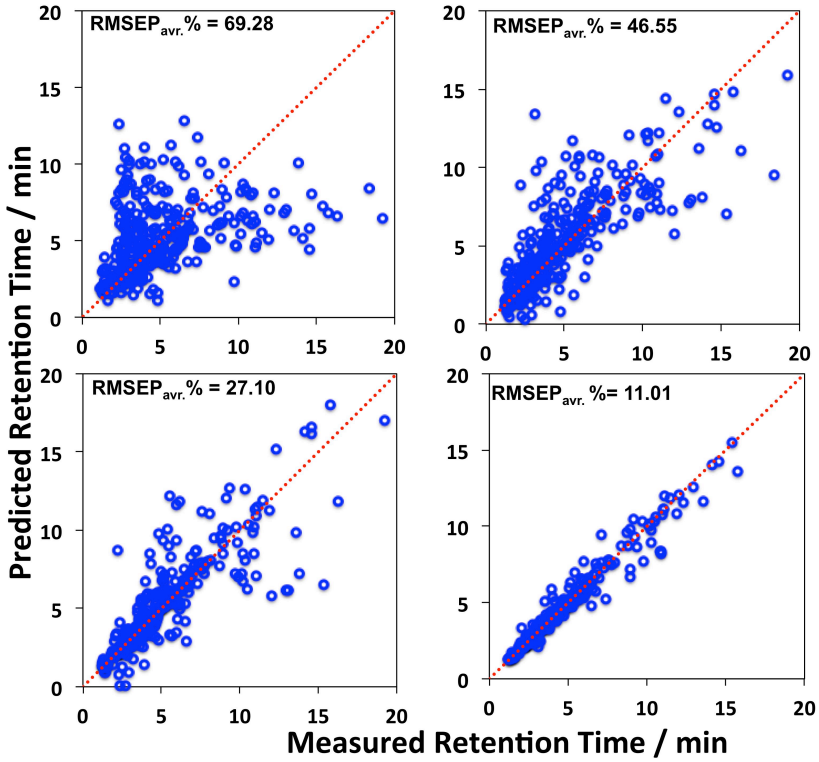

Figure 5.6. Predictive ability of (A) diverse, (B) global, (C) TS-based, and (D) dual-filtering based GA-PLS models for an external test set over five different HILIC systems. $RMSEP_{avr.}$ is the average value of RMSEP of test analytes over five HILIC stationary phases.

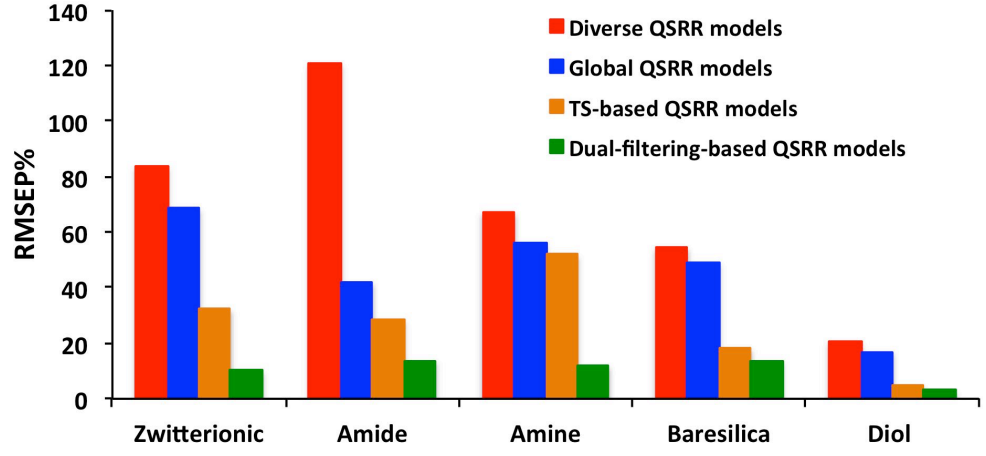

Figure 5.7. Comparison of the predictive ability of diverse, global, TS-based and dual-filtering-based GA-PLS models for an external test set over five different HILIC systems.

5.3.4 Relationship between molecular descriptors and the HILIC mechanism

The partitioning of the analytes between an immobilized water-enriched layer on a hydrophilic stationary phase and the relatively hydrophobic eluent of the bulk mobile phase has been recognised as a significant factor in the HILIC mechanism [37]. Besides partitioning, electrostatic interactions and hydrogen bonding have also been reported to contribute significantly to retention behaviour in HILIC systems [13, 38, 39]. However, the retention mechanism occurring in HILIC is acknowledged to be complex and a complete theoretical description has not yet been discovered [40, 41]. In this study, by analysis of the descriptors appearing in the proposed dual-filtering strategy for each stationary phase (Tables 5.10 and 5.11), some insight can be gained into the factors that influence the retention mechanism in these systems. Although the relevance of some descriptors is still not completely clear in terms of their significance to the chromatographic process, other descriptors are clearly linked with chemical properties that are relevant to the HILIC experimental system.

The importance of analyte partitioning in the HILIC mechanism is underlined by the structural descriptors which were most strongly correlated to t_R for each of the studied columns (see Table 5.10). The descriptor MLOGP (Moriguchi octanol-water partition coefficient (logP)) is selected in the dual-filtering process for the bare silica, amine and zwitterionic systems, while Hy (hydrophilic factor) is selected in the amide and diol systems. HOMT [42] and HATS2p [43, 44] belong to geometrical descriptors that capture information on molecular size and shape combined with the information on electronic properties. 2D autocorrelation descriptors [45] weighted by I-state (ATSC1s and ATSC4s) and van der Waals volume

(MATS1v) represent the role of atomic size and electronic properties in the retention behaviour of studied analytes in HILIC systems.

5.3.5 Conclusions

In this study, a novel compound-filtering-based QSRR modelling strategy that comprises a combination of Tanimoto similarity (TS) ranking and t_R similarity searching prior to GA-PLS modelling was demonstrated using HILIC data examples. The GA-PLS model based on a dual-filtering strategy permits a robust and accurate prediction of the retention times of test pharmaceuticals compared to global, diverse and TS-based QSRR models over a wide range of HILIC stationary phases. These results suggest that a QSRR model can reliably predict a test analyte's retention if the test analyte is sufficiently similar in structure and retention time to the group of compounds used to generate that QSRR model. The proposed method also contains sufficient information to enhance the understanding of the HILIC retention mechanism. Application of the proposed methodology in HILIC method development to identifying initial target columns and chromatographic conditions for unknown sample tests is the goal of future research.

5.4 References

[1] F.G. Vogt, A.S. Kord, Development of quality-by-design analytical methods, J. Pharm. Sci., 100 (2011) 797-812.

[2] E. Rozet, P. Lebrun, P. Hubert, B. Debrus, B. Boulanger, Design spaces for analytical methods, TrAC, Trends Anal. Chem., 42 (2013) 157-167.

[3] M. Jovanovic, T. Rakic, A. Tumpa, B. Jancic Stojanovic, Quality by Design approach in the development of hydrophilic interaction liquid chromatographic method for the analysis of iohexol and its impurities, J. Pharm. Biomed. Anal., 110 (2015) 42-48.

[4] S. Orlandini, B. Pasquini, M. Del Bubba, S. Pinzauti, S. Furlanetto, Quality by design in the chiral separation strategy for the determination of enantiomeric impurities: development of a capillary electrophoresis method based on dual cyclodextrin systems for the analysis of levosulpiride, J. Chromatogr. A, 1380 (2015) 177-185.
[5] K. Muteki, J.E. Morgado, G.L. Reid, J. Wang, G. Xue, F.W. Riley, J.W. Harwood, D.T. Fortin, I.J. Miller, Quantitative structure retention relationship models in an analytical quality by design framework: simultaneously accounting for compound properties, mobile-phase conditions, and stationary-phase properties, Ind. Eng. Chem. Res., 52 (2013) 12269-12284.
[6] R. Put, Y. Vander Heyden, Review on modelling aspects in reversed-phase liquid chromatographic quantitative structure-retention relationships, Anal. Chim. Acta, 602 (2007) 164-172.
[7] Y. Guo, Recent progress in the fundamental understanding of hydrophilic interaction chromatography (HILIC), Analyst, 140 (2015) 6452-6466.
[8] L.R. Snyder, J.W. Dolan, P.W. Carr, The hydrophobic-subtraction model of reversed-phase column selectivity, J. Chromatogr. A, 1060 (2004) 77-116.
[9] R. Kaliszan, QSRR: quantitative structure-(chromatographic) retention relationships, Chem. Rev., 107 (2007) 3212-3246.
[10] D.J. Creek, A. Jankevics, R. Breitling, D.G. Watson, M.P. Barrett, K.E. Burgess, Toward global metabolomics analysis with hydrophilic interaction liquid chromatography-mass spectrometry: improved metabolite identification by retention time prediction, Anal. Chem., 83 (2011) 8703-8710.
[11] N.S. Quiming, N.L. Denola, I. Ueta, Y. Saito, S. Tatematsu, K. Jinno, Retention prediction of adrenoreceptor agonists and antagonists on a diol

column in hydrophilic interaction chromatography, Anal. Chim. Acta, 598 (2007) 41-50.
[12] T. Baczek, R. Kaliszan, Predictions of peptides' retention times in reversed-phase liquid chromatography as a new supportive tool to improve protein identification in proteomics, Proteomics, 9 (2009) 835-847.
[13] G. Schuster, W. Lindner, Comparative characterization of hydrophilic interaction liquid chromatography columns by linear solvation energy relationships, J. Chromatogr. A, 1273 (2013) 73-94.
[14] R.I. Chirita, C. West, S. Zubrzycki, A.L. Finaru, C. Elfakir, Investigations on the chromatographic behaviour of zwitterionic stationary phases used in hydrophilic interaction chromatography, J. Chromatogr. A, 1218 (2011) 5939-5963.
[15] T. Körtvélyesi, M. Görgényi, K. Héberger, Correlation between retention indices and quantum-chemical descriptors of ketones and aldehydes on stationary phases of different polarity, Analytica chimica acta, 428 (2001) 73-82.
[16] M. Di Tullio, C. Maccallini, A. Ammazzalorso, L. Giampietro, R. Amoroso, B. De Filippis, M. Fantacuzzi, P. Wiczling, R. Kaliszan, QSAR, QSPR and QSRR in terms of 3-D-MoRSE descriptors for in silico screening of clofibric acid analogues, Mol Inform, 31 (2012) 453-458.
[17] G. Carlucci, A.A. D'Archivio, M.A. Maggi, P. Mazzeo, F. Ruggieri, Investigation of retention behaviour of non-steroidal anti-inflammatory drugs in high-performance liquid chromatography by using quantitative structure-retention relationships, Analytica chimica acta, 601 (2007) 68-76.
[18] G.M. Randazzo, D. Tonoli, S. Hambye, D. Guillarme, F. Jeanneret, A. Nurisso, L. Goracci, J. Boccard, S. Rudaz, Prediction of retention time in reversed-phase liquid chromatography as a tool for steroid identification, Analytica chimica acta, 916 (2016) 8-16.

[19] M. Goodarzi, R. Jensen, Y. Vander Heyden, QSRR modeling for diverse drugs using different feature selection methods coupled with linear and nonlinear regressions, J. Chromatogr. B Analyt. Technol. Biomed. Life Sci., 910 (2012) 84-94.
[20] M. Talebi, G. Schuster, R.A. Shellie, R. Szucs, P.R. Haddad, Performance comparison of partial least squares-related variable selection methods for quantitative structure retention relationships modelling of retention times in reversed-phase liquid chromatography, J. Chromatogr. A, 1424 (2015) 69-76.
[21] P. Zuvela, J.J. Liu, K. Macur, T. Baczek, Molecular descriptor subset selection in theoretical peptide quantitative structure-retention relationship model development using nature-inspired optimization algorithms, Anal. Chem., 87 (2015) 9876-9883.
[22] J.M. A., M.G. M., In Concepts and Applications of Molecular Similarity, John Wiley & Sons, New York, 1990.
[23] R.P. Sheridan, B.P. Feuston, V.N. Maiorov, S.K. Kearsley, Similarity to molecules in the training set is a good discriminator for prediction accuracy in QSAR, J. Chem. Inf. Comput. Sci., 44 (2004) 1912-1928.
[24] H. Yuan, Y. Wang, Y. Cheng, Local and global quantitative structure-activity relationship modeling and prediction for the baseline toxicity, J. Chem. Inf. Model., 47 (2007) 159-169.
[25] C.A. Bergstrom, C.M. Wassvik, U. Norinder, K. Luthman, P. Artursson, Global and local computational models for aqueous solubility prediction of drug-like molecules, J. Chem. Inf. Comput. Sci., 44 (2004) 1477-1488.
[26] H. Zhang, H.Y. Ando, L. Chen, P.H. Lee, On-the-fly selection of a training set for aqueous solubility prediction, Mol. Pharm., 4 (2007) 489-497.

[27] L. He, P.C. Jurs, Assessing the reliability of a QSAR model's predictions, J. Mol. Graph. Model., 23 (2005) 503-523.
[28] A.G. Maldonado, J.P. Doucet, M. Petitjean, B.T. Fan, Molecular similarity and diversity in chemoinformatics: from theory to applications, Mol. Divers., 10 (2006) 39-79.
[29] C. Wang, M.J. Skibic, R.E. Higgs, I.A. Watson, H. Bui, J. Wang, J.M. Cintron, Evaluating the performances of quantitative structure-retention relationship models with different sets of molecular descriptors and databases for high-performance liquid chromatography predictions, J. Chromatogr. A, 1216 (2009) 5030-5038.
[30] M. Talebi, S.H. Park, M. Taraji, Y. Wen, R.I.J. Amos, P.R. Haddad, R.A. Shellie, R. Szucs, C.A. Pohl, J.W. Dolan, Retention time prediction based on molecular structure in pharmaceutical method development: A perspective, LCGC, 34 550–558.
[31] Holliday, Grouping of coefficients for the calculation of inter-molecular similarity and aissimilarity using 2D fragment bit-strings, Comb. Chem. High Throughput Screening, 5 (2002).
[32] M. Taraji, R. Amos, M. Talebi, R. Szucs, C.A. Pohl, J.W. Dolan, P.R. Haddad, Rapid method development in hydrophilic interaction liquid chromatography for pharmaceutical analysis using a combination of quantitative structure–retention relationships and design of experiments, Anal. Chem., 89 (2017) 1870–1878.
[33] M. Taraji, Unreported work.
[34] E. Tyteca, M. Talebi, R. Amos, S.H. Park, M. Taraji, Y. Wen, R. Szucs, C.A. Pohl, J.W. Dolan, P.R. Haddad, J. Chromatogr. A, 1486 (2016) 50-58.
[35] in, For details on the hashed ChemAxon fingerprint developed by ChemAxon see

https://docs.chemaxon.com/display/CD/Chemical+Hashed+Fingerprint [accessed January 2016].

[36] W.A. Warr, Representation of chemical structures, Wiley Interdiscip. Rev.: Comput. Mol. Sci., 1 (2011) 557-579.

[37] A.J. Alpert, Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds, J. Chromatogr. A, 499 (1990) 177-196.

[38] A.E. Karatapanis, Y.C. Fiamegos, C.D. Stalikas, A revisit to the retention mechanism of hydrophilic interaction liquid chromatography using model organic compounds, J. Chromatogr. A, 1218 (2011) 2871-2879.

[39] Y. Guo, S. Gaiki, Retention and selectivity of stationary phases for hydrophilic interaction chromatography, J. Chromatogr. A, 1218 (2011) 5920-5938.

[40] B. Buszewski, S. Noga, Hydrophilic interaction liquid chromatography (HILIC)--a powerful separation technique, Anal. Bioanal. Chem., 402 (2012) 231-247.

[41] P. Jandera, Stationary phases for hydrophilic interaction chromatography, their characterization and implementation into multidimensional chromatography concepts, J. Sep. Sci., 31 (2008) 1421-1437.

[42] F. Feixas, E. Matito, J. Poater, M. Sola, On the performance of some aromaticity indices: a critical assessment using a test set, J. Comput. Chem., 29 (2008) 1543-1554.

[43] V. Consonni, R. Todeschini, M. Pavan, Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors. 1. Theory of the Novel 3D Molecular Descriptors, J. Chem. Inf. Model., 42 (2002) 682-692.

[44] V. Consonni, R. Todeschini, M. Pavan, P. Gramatica, Structure/Response Correlations and Similarity/Diversity Analysis by

GETAWAY Descriptors. 2. Application of the Novel 3D Molecular Descriptors to QSAR/QSPR Studies, J. Chem. Inf. Model., 42 (2002) 693-705.

[45] R. Todeschini, V. Consonni, Molecular Descriptors for Chemoinformatics, Wiley-WCH, Weinheim, 2009.

6 General conclusions

The development of computer-assisted approaches capable of accurate prediction of the retention behaviour of analytes, leading to optimisation of chromatographic performance, is a major goal for method development in chromatography [1]. Statistically derived quantitative structure-retention relationships (QSRRs) represent a quite popular approach to retention prediction [2]. A QSRR shows the relationships between chromatographic parameters, such as the retention as the dependent variable, and parameters describing analytes and columns for a given set of test molecules and columns.

Hydrophilic interaction chromatography (HILIC) [3] using a polar sorbent in combination with a hydro-organic mobile phase, provides an approach for the effective separation and quantitative determination of small polar compounds. Recently, HILIC has been successfully applied for the analyses of a wide range of small polar compounds, including drugs, toxins, plant extracts, and other compounds important to food and pharmaceutical industries [4, 5]. The detailed retention mechanism applicable in HILIC is still under some discussion and for this reason, method development in HILIC is difficult.

This thesis describes the development of retention prediction models for a variety of differently structured pharmaceutical compounds and commercially available stationary phases used in the HILIC mode.

QSRR models were developed to predict the retention times of analytes on five HILIC stationary phases (bare silica, amine, amide, diol and zwitterionic), with a view to selecting the most suitable stationary phase(s) for the separation of these analytes. The study was conducted using β-adrenergic agonists as target analytes. Molecular descriptors were calculated

based only on chemical structures optimised using density functional theory. A genetic algorithm (GA) was then used to select the most relevant molecular descriptors and these were used to build a retention model for each stationary phase using partial least squares (PLS) regression. This model was then used to predict the retention of the test set of target analytes. This process created an optimised descriptor set which enhanced the reliability of the developed QSRR models. Finally, the QSRR models developed in the work were utilised to provide some insight into the separation mechanisms operating in the HILIC mode. Three performance criteria – mean absolute error (MAE), root mean square error of prediction scaled to retention time (RMSEP), and the number of selected descriptors, were used to evaluate the developed models when applied to an external test set of β-adrenergic agonists and showed highly predictive abilities. RMSEP values of 4.88-11.12% were recorded. Validation was performed through Y-randomization and chemical domain applicability, from which it was evident that the developed optimised GA-PLS models were robust. The high levels of accuracy, reliability and applicability of the models were to a large extent due to the optimisation of the GA descriptor set and the presence of relevant structural and geometric molecular descriptors, together with descriptors based on important physicochemical properties, which establish a strong connection between retention time and meaningful chemical properties.

The present strategy holds great promise for broader screening of HILIC stationary phases for a desired separation, as well as for acquisition of information about molecular mechanisms of separation under chromatographic conditions.

Next, a design-of-experiment (DoE) model was developed, able to describe the retention times of a mixture of pharmaceutical compounds in

HILIC under all possible combinations of acetonitrile content, salt concentration, and mobile-phase pH with R^2 >0.95. Further, a QSRR model was developed to predict retention times for new analytes, based only on their chemical structures, with a RMSEP as low as 0.81%. A compound classification based on the concept of similarity was applied *prior to* QSRR modelling. Finally, a combined QSRR-DoE approach was utilised to propose an optimal design space in a quality-by-design (QbD) workflow to facilitate the HILIC method development. The mathematical QSRR-DoE model was shown to be highly predictive when applied to an independent test set of unseen compounds under unseen mobile phase conditions with a RMSEP value of 5.83%. The QSRR-DoE computed retention time of pharmaceutical test analytes and subsequently calculated separation selectivity was used to optimise the chromatographic conditions for efficient separation of targets. A Monte Carlo simulation was performed to evaluate the risk of uncertainty in the model's prediction, and to define the design space where the desired quality criterion was met. Experimental realization of peak selectivity between targets under the selected optimal working conditions confirmed the theoretical predictions. These results demonstrate how discovery of optimal conditions for the separation of new analytes can be accelerated by the combination of high-throughput theoretical and experimental methods.

The development of QSRRs with a sufficient accuracy to support high performance liquid chromatography (HPLC) method development is still a major issue [1, 6]. To tackle this challenge, this thesis has presented a novel QSRR methodology based on a dual filtering strategy which combined Tanimoto similarity (TS) searching as the primary filter and retention time (t_R) similarity clustering as the secondary filter, using a database of pharmaceutical compound retention times collected over a wide range of HILIC systems. To employ t_R similarity filtering, correlation to a molecular

descriptor was used as a measure of retention time. For the retention time of a compound to be modelled, a relationship between experimental chromatographic data and various molecular descriptors was calculated using a GA-PLS regression. The proposed dual-filtering-based QSRR model significantly improved the retention time predictability compared to the diverse, global and TS-based QSRR models, with an average RMSEP of 11.01% over five different HILIC stationary phases. Interpretation of the molecular descriptor correlation strategy revealed that particular trends for the HILIC mechanism could be captured using the proposed dual filtering technique.

References

[1] T. Bolanča, Š. Ukić, M. Novak, M. Rogošić, Computer assisted method development in liquid chromatography, Croat. Chem. Acta, 87 (2014) 111-122.

[2] R. Kaliszan, QSRR: quantitative structure-(chromatographic) retention relationships, Chem. Rev., 107 (2007) 3212-3246.

[3] A.J. Alpert, Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds, J. Chromatogr. A, 499 (1990) 177-196.

[4] P. Hemström, K. Irgum, Hydrophilic interaction chromatography, J. Sep. Sci., 29 (2006) 1784-1821.

[5] B. Buszewski, S. Noga, Hydrophilic interaction liquid chromatography (HILIC)--a powerful separation technique, Anal. Bioanal. Chem., 402 (2012) 231-247.

[6] K. Muteki, J.E. Morgado, G.L. Reid, J. Wang, G. Xue, F.W. Riley, J.W. Harwood, D.T. Fortin, I.J. Miller, Quantitative structure retention relationship models in an analytical quality by design framework: simultaneously accounting for compound properties, mobile-phase

conditions, and stationary-phase properties, Ind. Eng. Chem. Res., 52 (2013) 12269-12284.